

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw mode:

Run on: October 24, 2003, 15:43:05 ; Search time 2765.41 Seconds
(without alignments)
14808.112 Million cell updates/sec

Title: US-09-830-902-1_COPY_100000_101000

Perfect score: 1001

Sequence: 1 aatataaaaatgcatacttt.....acctgattctgttctttt 1001

Scoring table: IDENTITY NJC
Gapop 10.0 , Gapext 1.0

Searched: 2898711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:

1: gb_ba:
2: gb_htg:
3: gb_in:
4: gb_om:
5: gb_ov:
6: gb_pat:
7: gb_ph:
8: gb_pi:
9: gb_pr:
10: gb_ro:
11: gb_sts:
12: gb_sy:
13: gb_un:
14: gb_vi:
15: em_ba:
16: em_fun:
17: em_hum:
18: em_in:
19: em_mu:
20: em_om:
21: em_or:
22: em_ov:
23: em_pat:
24: em_ph:
25: em_pi:
26: em_ro:
27: em_sts:
28: em_un:
29: em_vi:
30: em_htg_hum:
31: em_htg_inv:
32: em_htg_other:
33: em_htg_mus:
34: em_htg_pin:
35: em_htg_rcd:
36: em_htg_mam:
37: em_htg_vrt:
38: em_sy:
39: em_htgo_hum:
40: em_htgo_mus:
41: em_htgo_other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1001	100.0	110000	6	AX093471	AX093471 Sequence
2	1001	100.0	110000	9	HSA246003	AJ246003 Homo sapi
3	1001	100.0	162692	9	CNS01DS8	AJ21659 BAC sequ
4	1001	100.0	185281	2	AC011232	AC011232 Homo sapi
5	403	40.3	3263	6	AX093472	AX093472 Sequence
6	403	40.3	3263	9	HSA246001	AJ246001 Homo sapi
7	403	40.3	5120	9	ABC29006	ABC29006 Homo sapi
8	200.6	20.0	172816	9	AC093899	AC093899 Homo sapi
9	199.9	20.0	195432	9	AC130403	AC130403 Homo sapi
10	199.8	20.0	203354	2	AC068752	AC068752 Homo sapi
11	197.8	19.8	163027	9	AC002549	AC002549 Homo sapi
12	197.8	19.8	163662	2	AC096508	AC096508 Homo sapi
13	197.8	19.8	191866	2	AC068676	AC068676 Homo sapi
14	196.8	19.7	119922	9	AC099804	AC099804 Homo sapi
15	195.8	19.6	46775	9	HSJ738A13	AL109801 Human DNA
16	195.6	19.5	182799	2	AC073526	AC073526 Homo sapi
17	195.4	19.5	130569	9	CNS08CB8	AL845552 Human chr
18	195.4	19.5	137506	9	AC004824	AC004824 Homo sapi
19	194.6	19.4	177300	9	AC096741	AC096741 Homo sapi
20	194.4	19.4	45373	9	AL158092	AL158092 Human DNA
21	194.4	19.4	173520	9	AL732423	AL732423 Human DNA
22	193.6	19.3	176498	9	AC003035	AC003035 Homo sapi
23	193.6	19.3	178908	9	AL157392	AL157392 Human DNA
24	193.6	19.3	189485	2	AC141268	AC141268 Homo sapi
25	193.4	19.3	81542	9	AC100830	AC100830 Homo sapi
26	193.4	19.3	111744	2	AC099686	AC099686 Homo sapi
27	193.4	19.3	150147	9	AC002543	AC002543 Homo sapi
28	193.4	19.3	202920	2	AC139392	AC139392 Rattus no
29	193.4	19.3	215829	2	AC131423	AC131423 Rattus no
30	193.2	19.3	82517	9	HS341D10	Z97985 Human DNA S
31	193	19.3	169476	9	CNS01DW7	AL136419 Human chr
32	192.8	19.3	140363	9	AC142331	AC142331 Pan trogi
33	192.8	19.3	172533	9	AC002381	AC002381 Human BAC
34	192.4	19.2	150489	2	AC069491	AC069491 Homo sapi
35	192.4	19.2	157574	9	AL592293	AL592293 Human DNA
36	192.4	19.2	160169	2	AC051664	AC051664 Homo sapi
37	192.4	19.2	163215	2	AC087451	AC087451 Homo sapi
38	192.4	19.2	166867	9	AP003733	AP003733 Homo sapi
39	192.4	19.2	178770	2	AC090725	AC090725 Homo sapi
40	192.4	19.2	181140	9	AP002793	AP002793 Homo sapi
41	192.4	19.2	205150	9	AC005274	AC005274 Homo sapi
42	192.2	19.2	145423	9	AP003086	AP003086 Homo sapi
43	192.2	19.2	161096	9	AC016257	AC016257 Homo sapi
44	192.2	19.2	167941	2	AC058804	AC058804 Homo sapi
45	192.2	19.2	185737	9	AC005468	AC005468 Homo sapi

ALIGNMENTS

RESULT 1
AX093471
LOCUS AX093471 110000 bp DNA linear PAT 30-MAR 2001
DEFINITION Sequence 1 from Patent WO0118198.
ACCESSION AX093471
VERSION AX093471.1 GI:13509911
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1
AUTHORS Weissbach, J. and Hazan, J.
TITLE Cloning, expression and characterization of the spg4 gene
responsible for the most frequent form of autosomal spastic

JOURNAL
parap'égia
Patent: WO 01.8198-A 1 15-MAR-2001;
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)

DB	100360	ATACCTATGAAACTAAACACAAAGCAATAATATATTAAGTATAGCATTTATTAAGGTTCTT	100419
QY	421	TGTGTACGAGATCAACATAGAAAATATATTTAAATGGCTGACATAATTTTCTAAGAAATAC	480
DB	100420	TGTGTACGAGATCAACATAGAAAATATATTTAAATGGCTGACATAATTTTCTAAGAAATAC	100479
QY	481	ATACACGTATATTTTATTAACATTTAAGAAACAGCAGCATCATTTAAATCCCATCAT	540
DB	100480	ATACACGTATATTTTATTAACATTTAAGAAACAGCAGCATCATTTAAATCCCATCAT	100539
QY	541	TGCTTAACCAACCATATACCTGTTGATCATTTTGTAATGTCAATGGCTTTTAAAAATCTAG	600
DB	100540	TGCTTAACCAACCATATACCTGTTGATCATTTTGTAATGTCAATGGCTTTTAAAAATCTAG	100599
QY	601	ATGAGAAATATTTCGATTATCTGACTTTCACCTGAATCCTTGAAAAAAATAAACGCAGCGTC	660
DB	100600	ATGAGAAATATTTCGATTATCTGACTTTCACCTGAATCCTTGAAAAAAATAAACGCAGCGTC	100659
QY	661	AGCCCTCAAACTTTAGAAAGCGTACATACGTTGGAACAAGGACTTTTGGAGATACCACTGTT	720
DB	100660	AGCCCTCAAACTTTAGAAAGCGTACATACGTTGGAACAAGGACTTTTGGAGATACCACTGTT	100719
QY	721	TAAGGAAATACCTTTGTMAACCTGCAGAACATTTTACTTTAAAGAGGAACACAAAGATCT	780
DB	100720	TAAGGAAATACCTTTGTMAACCTGCAGAACATTTTACTTTAAAGAGGAACACAAAGATCT	100779
QY	781	TCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTATAGAGTCTTAC	840
DB	100780	TCAATGAACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTATAGAGTCTTAC	100839
QY	841	ATATTTGTGCACCAAACTTTGAAGATGAACCCAGAAAACAGACTTAAACAAAATATACAATG	900
DB	100840	ATATTTGTGCACCAAACTTTGAAGATGAACCCAGAAAACAGACTTAAACAAAATATACAATG	100899
QY	901	CAAAATGTAATTTTGTGTTTAAAGCCCTTGCCCTTGATGGTCACAGTTATCCCAATGGAC	960
DB	100900	CAAAATGTAATTTTGTGTTTAAAGCCCTTGCCCTTGATGGTCACAGTTATCCCAATGGAC	100959
QY	961	ACTAAGTTTAGAGCAACAACAAACCTGATTCCTGGTCTTCTTT	1001
DB	100960	ACTAAGTTTAGAGCAACAACAAACCTGATTCCTGGTCTTCTTT	101000

Query Match	100.0%	Score 1001	DB 6	Length 110000
Best Local Similarity	100.3%	Pred. No. 9.7e-181		
Matches 1001	Conservative	0	Mismatches	0
			Indels	0
			Gaps	0
QY	1	AATATAAAATGCATACCTTAGGCTGGGCAAGGTGGCTCACGCTCTGTAAATCCAGCAGCT	60	
DB	100050			
QY	61	TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTAGGACCAACACTGGCCATCA	120	
DB	100060			
QY	121	TGGCGAAACCGCTGCTCTACTAAAAATACAAAAATTAGCTGGGCGATGGAGGCATGTGCCT	180	
DB	100120			
QY	181	GTAATCCAGCTACTCGGGAGGCTGAGACAGAAAGTAATTCCTTGAACCTTGGAGGGGAGG	240	
DB	100180			
QY	241	TTGCATATCTGAGTGGTGAATGTGATCTCTTTTCTCTTTGCTGTATTTTTGAACCT	300	
DB	100240			
QY	301	TTTCTATAAATGATGTGTTTTTGTTTTCTATAATGGAAAAAATATATATGCTTTCAAAATGTTA	360	
DB	100300			
QY	361	ATACCTATGAACTAAACACAACTAATAAATATATAGTATAGCACTTATTAAGGTTTCT	420	
DB	100359			

BASE COUNT 30522 a 21640 c 22817 g 34921 l
ORIGIN

TITLE	Spectrum of SP34 mutations in autosomal dominant spastic paraplegia	
JOURNAL	Hum. Mol. Genet. 9 (4), 637-644 (2000)	
MEDLINE	20164302	
PUBMED	10699187	
REFERENCE	3 (bases 1 to 110000)	
AUTHORS	Genoscope.	
TITLE	Direct Submission	
JOURNAL	Submitted (17-JUN-1999) Genoscope, Genoscope - Centre National de	
	Sequencage, BP 191, Evry 91006, FRANCE	
COMMENT	E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr. The sequence is the result of the assembly of 2 BAC clones: R-336p14 and 563N4, respectively from RPCI-11 and CITE_97a_SK3 library.	
FEATURES	Location/Qualifiers	
source	1. .110000	
	/organism="Homo sapiens"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:9606"	
	/chromosome="2"	
	/map="2p21"	
	/clone="R-336p14"	
	/clone_lib="RPCI-11"	
	/note="BAC R-336p14 contains Spast gene: exon 1 to exon 3 ; BAC 563N4 contains Spast gene: exon 4 to exon 17"	
gene	9312..102009	
	/gene="Spast"	
mRNA	9312..102009	
	/gene="Spast"	
	/product="spastin protein"	
exon	9312..10471	
	/gene="Spast"	
	/number=1	
5' UTR	9312..10056	
	/gene="Spast"	
CDS	join(10057..10471,33719..33805,35749..35832,45023..45118, 60864..61051,61928..62061,62345..62438,73174..73248, 74634..74705,82789..82864,83103..83194,83335..83414, 88130..88172,89562..89641,91163..91233,93444..93484, 103600..100722)	
	/gene="Spast"	
	/codon_start=1	
	/product="spastin protein"	
	/protein_id="CAB60208.1"	
	/db_xref="GI:6273493"	
	/translation="MNSPGGRGKKKGSGASNPVPRP??PCLAPAPPAGAPPPEP PKKNLYFSYPLFVGFALLRQVAFHLGLFVWLCOFESRALMAKRSSGAPAPASA SAPAPVGGAEAEVRVFKQAFYISIALRIDEDEKAGQKEQAVENWYKKGIELEKGI AVIVTGGEGQECERARRLQAKMTNLMVAKORLQJLEKMPVLPPFSKSTQDYNDSTN ACRNHLQESGAVPKRKDPLTH?SNSLPRSKTVMTKTSAGLSGHHRAPSYSGLSMVS GKQSGSPAP?THKGT?PKTNRTNKPSTPTATPKKKKDLKNFNVDNLANLIMNEIVD NGTAVFDDIAGQDLAKALQEIIVLPSLRPELFTGLRAPAGLLLFSPGNGKTMJA KAAAESNATFN?SAAASLTSKYVGEKLVRLPAVARELOPSIIFI?DEVDSLLCER REGHDCASRRJX?EFLIEFDGVSAGDDRVLVYGATNRPQEDAVLRRPIKRVVYSL PNEETRLLLKNLJLCKQGSPLIQKELAQLRMTDGYSGSD?TALAKDAALGPIRELKP EQVKNMSASEMRN?RLSDFTESLKKIKRSVSPQTLEAYIRWNKQFGDTTV"	
intron	10472..33718	
	/gene="Spast"	
	/number=1	
exon	33719..33805	
	/gene="Spast"	
	/number=2	
intron	33806..35748	
	/gene="Spast"	
	/number=2	
exon	35749..35932	
	/gene="Spast"	
	/number=3	
intron	35933..45022	
	/gene="Spast"	
	/number=3	
exon	45023..45118	
	/gene="Spast"	
	/number=4	

intron	45119..60863	
	/gene="Spast"	
	/number=4	
exon	60864..61051	
	/gene="Spast"	
	/number=5	
intron	61052..61927	
	/gene="Spast"	
	/number=5	
exon	61928..62061	
	/gene="Spast"	
	/number=6	
intron	62062..62344	
	/gene="Spast"	
	/number=6	
exon	62345..62438	
	/gene="Spast"	
	/number=7	
intron	62439..73173	
	/gene="Spast"	
	/number=7	
exon	73174..73248	
	/gene="Spast"	
	/number=8	
intron	73249..74633	
	/gene="Spast"	
	/number=8	
exon	74634..74705	
	/gene="Spast"	
	/number=9	
intron	74706..82788	
	/gene="Spast"	
	/number=9	
exon	82789..82864	
	/gene="Spast"	
	/number=10	
intron	82865..83102	
	/gene="Spast"	
	/number=10	
exon	83103..83194	
	/gene="Spast"	
	/number=11	
intron	83195..83334	
	/gene="Spast"	
	/number=11	
exon	83335..83414	
	/gene="Spast"	
	/number=12	
intron	83415..88129	
	/gene="Spast"	
	/number=12	
exon	88130..88172	
	/gene="Spast"	
	/number=13	
intron	88173..89561	
	/gene="Spast"	
	/number=13	
exon	89562..89641	
	/gene="Spast"	
	/number=14	
intron	89642..91162	
	/gene="Spast"	
	/number=14	
exon	91163..91233	
	/gene="Spast"	
	/number=15	
intron	91234..93443	
	/gene="Spast"	
	/number=15	
exon	93444..93484	
	/gene="Spast"	
	/number=16	
intron	93485..100599	

```
/gene="Spast"
/number=16
100600..102009
/gene="Spast"
/number=17
100721..102009
/gene="Spast"
/number=18
101973..101978
/gene="Spast"
polya_signal
101973..101978
BASE COUNT 30622 a 21840 c 22817 g 34921 t
ORIGIN

Query Match      100.0%;   Score 1001;   DB 9;   Length 110000;
Best Local Similarity 100.0%;   Pred. No. 9.4e-181;
Matches 1001;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

QY 1 AATATAAAATGATACATCTTTAGGCTGGGCAAGTGGCTCAGCTCTGTATATCCAGCACTT 60
DB 100000 AATATAAAATGATACATCTTTAGGCTGGGCAAGTGGCTCAGCTCTGTATATCCAGCACTT 100059
QY 61 TGGGAGGCCAAGTGGCGGATCACCTGAGGTAGGATCAGGATCAGGACCAACCTGGCCATCA 120
DB 100060 TGGGAGGCCAAGTGGCGGATCACCTGAGGTAGGATCAGGATCAGGACCAACCTGGCCATCA 100119
QY 121 TGGGGAACCCCTGCTCTACTAATAAATACAAATATAGCTGGGCATGAGGCACTGTGCTT 180
DB 100120 TGGGGAACCCCTGCTCTACTAATAAATACAAATATAGCTGGGCATGAGGCACTGTGCTT 100179
QY 181 GTAATCCCACTACTCGGAGGCTGAGACAGAAGAAATGCTTGAACCTTGGAGGGGAGG 240
DB 100180 GTAATCCCACTACTCGGAGGCTGAGACAGAAGAAATGCTTGAACCTTGGAGGGGAGG 100239
QY 241 TTGCATATCTGACTGGTGAATTTGTGATTCCTTTTCTCTCTTGTCTGCTGATTTTGGACT 300
DB 100240 TTGCATATCTGACTGGTGAATTTGTGATTCCTTTTCTCTCTTGTCTGCTGATTTTGGACT 100299
QY 301 TTCTATAAATGATTGTGTTTGTGTTTATATATGGAAATATATGCTTCAATGTTA 360
DB 100300 TTCTATAAATGATTGTGTTTGTGTTTATATGGAATATATGCTTCAATGTTA 100359
QY 361 ATACCTATGAACATAACACACAGTANTAAATATATAGTATAGCATTTATTAAGTTTCT 420
DB 100360 ATACCTATGAACATAACACACAGTANTAAATATATAGTATAGCATTTATTAAGTTTCT 100419
QY 421 TGTGTAGCAGATCAACATAGAAATATATTTAAATGGCTGACATATTTTCTAAGAAATAC 480
DB 100420 TGTGTAGCAGATCAACATAGAAATATATTTAAATGGCTGACATATTTTCTAAGAAATAC 100479
QY 481 ATACACGTATATTTTATAACATTAAGAAACAGCAGCATCATTTACTTTAATCCATCAT 540
DB 100480 ATACACGTATATTTTATAACATTAAGAAACAGCAGCATCATTTACTTTAATCCATCAT 100539
QY 541 TCGTTAACCCACCATATACCTGTTGATCATTTTGTATTGTCATGTGCTTTTAAATATCTAG 600
DB 100540 TCGTTAACCCACCATATACCTGTTGATCATTTTGTATTGTCATGTGCTTTTAAATATCTAG 100599
QY 601 ATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAATAAATAAAGCGCAGCTC 660
DB 100600 ATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAATAAATAAAGCGCAGCTC 100659
QY 661 AGCCCTCAAACTTTAGAGCGGTACATACGTTTGGAAACAAGGACTTTGGAGATACCACTGT 720
DB 100660 AGCCCTCAAACTTTAGAGCGGTACATACGTTTGGAAACAAGGACTTTGGAGATACCACTGT 100719
QY 721 TAAGAAATACCTTTGTAAACCTGACAGAAATTTTACTTTAAAGAGGAAACACAGATCT 780
DB 100720 TAAGAAATACCTTTGTAAACCTGACAGAAATTTTACTTTAAAGAGGAAACACAGATCT 100779
QY 781 TCAATGAACGTCATCGGCTACAGAAACAGCCTTAAGTTTACAGGACTTTTACAGTCTTAC 840
DB 100780 TCAATGAACGTCATCGGCTACAGAAACAGCCTTAAGTTTACAGGACTTTTACAGTCTTAC 100839
QY 841 ATATTTGTGACCAAACTTGAAGATGAACCAAGAAACAGACTTAAACAAATATACATG 900
```


64473 ANTATAAAATGCATACCTTTAGGCTGGGCAAGTGGCTCAGCTGTGTAAATCCACGACCTT 64414
QY 61 TGGAGGCGCAAGGTGGGCGATCAGCTGAGGTTAGGAGTTTCCAGGACCAACCTGGCCATCA 120
Db 64413 TGGAGGCGCAAGGTGGGCGATCAGCTGAGGTTAGGAGTTTCCAGGACCAACCTGGCCATCA 64354
QY 121 TGGCGAAACCCCTGTCTCTACTAAAAATACAAAAATFASCTGGGCTGGAGGCAATGTGCGT 180
Db 64353 TGGCGAAACCCCTGTCTCTACTAAAAATACAAAAATFASCTGGGCTGGAGGCAATGTGCGT 64294
QY 191 GTAAATCCCAAGCTACTCGGAGGCTGAGACAGAGAAATFAGCTTGAACCTTGGAGGGGAGG 240
Db 64293 GTAAATCCCAAGCTACTCGGAGGCTGAGACAGAGAAATFAGCTTGAACCTTGGAGGGGAGG 64234
QY 241 TTGCATATCTGAGTGGTGAATTTGGATCTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 300
Db 64233 TTGCATATCTGAGTGGTGAATTTGGATCTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCT 64174
QY 301 TTCTATAAATGATCTGTGTTTGTGTTTATATTGGAAAAATATTATGCTTTTCAATGTTA 360
Db 64173 TTCTATAAATGATCTGTGTTTGTGTTTATATTGGAAAAATATTATGCTTTTCAATGTTA 64114
QY 361 ATACCTATGAACTAAACACAGTAATAAATATATTACTATAGCATTTATTAGGTTCT 420
Db 64113 ATACCTATGAACTAAACACAGTAATAAATATATTACTATAGCATTTATTAGGTTCT 64054
QY 421 TGTGTAGCAGATCAACATAGAAATATATTAAATGCTGACATATTTTCTAGAAATAC 480
Db 64053 TGTGTAGCAGATCAACATAGAAATATATTAAATGCTGACATATTTTCTAGAAATAC 63994
QY 481 ATACAGCTATATTTTAAACATTAAGAAACAGCAGCATCTTACTTAAATGATCAT 540
Db 63993 ATACAGCTATATTTTAAACATTAAGAAACAGCAGCATCTTACTTAAATGATCAT 63934
QY 541 TCGTTAAACACCATATACCTGTGTGATCATTTGTAGTGTGATGCTTTTAAATCTAG 600
Db 63933 TCGTTAAACACCATATACCTGTGTGATCATTTGTAGTGTGATGCTTTTAAATCTAG 63874
QY 601 ATGAGAAATATTGATTTATCTGACTTCACTGAATCCTTGAATAAATAAAGCAGCGTC 660
Db 63873 ATGAGAAATATTGATTTATCTGACTTCACTGAATCCTTGAATAAATAAAGCAGCGTC 63814
QY 661 AGCCCTCAAACTTTAGAGGCTTACATACGTTGGAACAGGACTTTGAGATACCACTGTT 720
Db 63813 AGCCCTCAAACTTTAGAGGCTTACATACGTTGGAACAGGACTTTGAGATACCACTGTT 63754
QY 721 TAAGGAATAGCTTTGTAAACCTGCAGAACATTTTACTTAAAGAGGAAACACAGATCT 780
Db 63753 TAAGGAATAGCTTTGTAAACCTGCAGAACATTTTACTTAAAGAGGAAACACAGATCT 63694
QY 781 TCAATGAACGTATCGGCTACAGAAACAGCTTAAGTTTACAGGACTTTTGTAGCTTTAC 840
Db 63693 TCAATGAACGTATCGGCTACAGAAACAGCTTAAGTTTACAGGACTTTTGTAGCTTTAC 63634
QY 841 ATATTTGTGACCAAACTTTGAGATGAACAGAAACAGACTTTAAACAAATATACAATG 900
Db 63633 ATATTTGTGACCAAACTTTGAGATGAACAGAAACAGACTTTAAACAAATATACAATG 63574
QY 901 CAAATGTAATTTTGTGTTTAAAGGCTTGCCTTGTATGCTCAGCTTATCCCAATGGAC 960
Db 63573 CAAATGTAATTTTGTGTTTAAAGGCTTGCCTTGTATGCTCAGCTTATCCCAATGGAC 63514
QY 961 ACTAAGTTAGAGCACAACAAACCTGATTTCTGCTCTCTTT 1001
Db 63513 ACTAAGTTAGAGCACAACAAACCTGATTTCTGCTCTCTTT 63473

RESULT 5
AX093472
LOCUS AX093472 3263 bp DNA linear PAT 10-MAR-2001
DEFINITION Sequence 2 from Patent WO0118198.
ACCESSION AX093472
VERSION AX093472.1 GI:113509912

KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE Weissenbach, J. and Hazan, J.
AUTHORS Cloning, expression and characterisation of the spg4 gene
TITLE responsible for the most frequent form of autosomal spastic
paraplegia
JOURNAL Patent: WO 0118198-A 2 15-MAR-2001;
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)
FEATURES Location/Qualifiers
source 1..3263
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 956 a 564 c 703 g 940 t
ORIGIN
Query Match 40.3%; Score 403; DB 6; Length 3263;
Best Local Similarity 100.0%; Pred. No. 5.7e-67;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 599 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAATAAATAAAGCGACCG 659
Db 1852 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAATAAATAAAGCGACCG 1911
QY 659 TCAGCCCTCAAACTTTAGAGGCTTACATACGTTGGAACAGGACTTTTGAGATACCACTG 718
Db 1912 TCAGCCCTCAAACTTTAGAGGCTTACATACGTTGGAACAGGACTTTTGAGATACCACTG 1972
QY 718 TTTAAGGAAATACCTTTGTAACCTTGAGAACATTTTACTTAAAGAGGAAACACAGAT 778
Db 1972 TTTAAGGAAATACCTTTGTAACCTTGAGAACATTTTACTTAAAGAGGAAACACAGAT 2031
QY 779 CTTCAATGAACGTCATCGCTACAGAAACAGCCTAAGTTTACAGGACTTTTAGAGTCTT 838
Db 2032 CTTCAATGAACGTCATCGCTACAGAAACAGCCTAAGTTTACAGGACTTTTAGAGTCTT 2092
QY 839 ACATATTTGTGACCAAACTTTGAAGATGAACAGAAACAGACTTAAACAAATATACAA 898
Db 2092 ACATATTTGTGACCAAACTTTGAAGATGAACAGAAACAGACTTAAACAAATATACAA 2151
QY 899 TGCAATATGTAATTTTGTGTTTAAAGGCTTGCCTTGTGATGCTTATCCCAATGG 958
Db 2152 TGCAATATGTAATTTTGTGTTTAAAGGCTTGCCTTGTGATGCTTATCCCAATGG 2211
QY 959 ACATTAAGTTAGAGCACAACAAACCTGATCTGCTCTCTTT 1001
Db 2212 ACATTAAGTTAGAGCACAACAAACCTGATCTGCTCTCTTT 2254
RESULT 6
HSA246001 3263 bp mRNA linear PRI 10-MAR-2000
LOCUS Homo sapiens mRNA for spastin protein (Spast gene).
DEFINITION
ACCESSION AJ246001
VERSION AJ246001.1 GI:6273493
KEYWORDS Spast gene; spastin protein; SPG4-linked hereditary spastic
paraplegia.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE Hazan, J., Fonknechten, N., Mavel, D., Paternotte, C., Samson, D.,
Artiguenave, F., Davoine, C.S., Craud, C., Durif, A., Wincker, P.,
Brottier, P., Cattolico, L., Barbe, V., Burgunder, J.M.,
Prud'homme, J.F., Brice, A., Fontaine, B., Hellig, R. and
Weissenbach, J.
AUTHORS Spastin, a new AAA protein, is altered in the most frequent form of
autosomal dominant spastic paraplegia

```
JOURNAL Nat. Genet. 23 (3), 296-303 (1999)
MEDLINE 20055425
PUBMED 10610178
REFERENCE 2
AUTHORS Fonknechten, N., Mavel, D., Byrne, P., Davoine, C., Cruaud, C.,
Boentsch, E., Samson, D., Coutinho, P., Hutchinson, M., McMonagle, P.,
Burgunder, J., Tartaglione, A., Heinzel, O., Feki, I., Douflet, T.,
Pattrey, N., Brice, A., Fontaine, B., Prud'homme, C., Weissensbach, J.,
Durr, A. and Hazan, C.
TITLE Spectrum of SP04 mutations in autosomal dominant spastic paraplegia
JOURNAL Hum. Mol. Genet. 9 (4), 637-644 (2000)
MEDLINE 20164302
PUBMED 10699187
REFERENCE 3 (bases 1 to 3263)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1999) Genoscope, Genoscope - Centre National de
Sequencage, BP 191, EVRY 91006, FRANCE
COMMENT E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr.
FEATURES
    source
        1. 3263
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /chromosome="2"
            /map="2p21"
        2. 3263
            /gene="Spast"
            /gene="Spast"
            /codon_start=1
            /product="spastin protein"
            /protein_id="CA660141.1"
            /db_xref="GI:6273491"
            /translation="MNSPGGRKKKGGSGASNPVPRPPPCJAPAPPAGPAPPES
            PHKRNLYFYSYPLFVGFALLRLVAFHLGLFVWLCQRFSRALMAKRSSGAPAPASA
            SAPAPVGGGEAEVRVVFHKQAFYISIALRIDEKAGQKEQAVENYKNGISELEKGI
            AVIVTQGEQGERARRLQAKMYNLVMAKORLQLEKMPVLPEKSKQTDVNDSTN
            ACNGHLSQESGAVPKRKOPLTHTSNSLPRSKTMKTSAGSGHHRAPSYSGLSMVS
            GVKGSQSPAPTHKGIPTKNTNKPSTPTTATRKKKOLKFNVDENLANLKNEIVD
            NGTAVKFDIAGQDLAKQALQEIIVLPSLRPELFTGLRAPARGLLFPGPGNGKTMIA
            KAVAAESNATFENSAASLTSKYVGEKLVRLFAVARELQFSIFIDEVLSLQCR
            REGHDAARLLKTEFLIEFDGVQSGADDRVLVVGATNRPOELDEAVLRRLKRVYVSL
            PNEETRLLLKNTLCKGSLTQKELACLARVTDGVSGSLTALAKDAAGPIRELKP
            EQVKNMSASEMRNRLSDFTESLKKIKRSVSPQLEAYIRWKKDFGDTTV"
            polyA_signal 3227..3232
            /gene="Spast"
BASE COUNT 956 a 564 c 703 g 940 t
ORIGIN
Query Match 40.3%; Score 403; DB 9; Length 3263;
Best Local Similarity 100.0%; Pred. No. 5.7e-67;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 599 AGATGAGAAATATCGATTATCGACTTCACCTGATCCCTTGAAAAAATAAAGCGGCG 658
Db 1852 AGATGAGAAATATCGATTATCGACTTCACCTGATCCCTTGAAAAAATAAAGCGGCG 1911
QY 659 TCAGCCCTCAAACCTTAGAAGCGTACATACCTTGAACACAGGACTTTGGAGATACCACTG 718
Db 1912 TCAGCCCTCAAACCTTAGAAGCGTACATACCTTGAACACAGGACTTTGGAGATACCACTG 1971
QY 719 TTTAGGAAATACCTTTGTAACCTGCAGAACATTTTACTTAAAGAGGGAACACAGAT 778
Db 1972 TTTAGGAAATACCTTTGTAACCTGCAGAACATTTTACTTAAAGAGGGAACACAGAT 2031
QY 779 CTTCAATGACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTAGAGTCTT 838
Db 2032 CTTCAATGACGTCATCGGCTACAGAAACAGCCTAAGTTTACAGGACTTTTAGAGTCTT 2091
QY 839 ACATATTTGTGCACCAAACTTGAAGATGAACCAAGAAACAGACTTTAAACAAATATACAA 898
Db 2092 ACATATTTGTGCACCAAACTTGAAGATGAACCAAGAAACAGACTTTAAACAAATATACAA 2151
```

```
QY 899 TSCAAATGTAATTTTGTGTTAAGGCGCTTGCTTGATGGTACAGATTATCCCAATGG 958
Db 2152 TSCAAATGTAATTTTGTGTTAAGGCGCTTGCTTGATGGTACAGATTATCCCAATGG 2211
QY 959 ACACCTAAGTTAGAGCAACAAACAAACCTGATCTGCTCTCTTT 1001
Db 2212 ACACCTAAGTTAGAGCAACAAACAAACCTGATCTGCTCTCTTT 2254

RESULT 7
LOCUS AB029006 5120 bp mRNA linear PRI 04-AUG-1999
DEFINITION Homo sapiens mRNA for KIAA1083 protein, complete cds.
ACCESSION AB029006
VERSION AB029006.1 GI:5689502
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS Kikuno, R., Nagase, T., Ishikawa, K., Hirose, M., Miyajima, N.,
Tanaka, A., Kotani, H., Nomura, N. and Ohara, O.
TITLE Prediction of the coding sequences of unidentified human genes.
XIV. The complete sequences of 100 new cDNA clones from brain which
code for large proteins in vitro
JOURNAL DNA Res. 6 (3), 197-205 (1999)
MEDLINE 99397452
PUBMED 10470851
REFERENCE 2 (bases 1 to 5120)
AUTHORS Chara, C., Nagase, T. and Kikuno, R.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1999) Osamu Ohara, Kazusa DNA Research Institute,
Laboratory of DNA Technology, Yana 1532-3, Kisarazu, Chiba
292-0812, Japan (E-mail:cdnaif@kazusa.or.jp, Tel:+81-438-52-3913,
Fax:+81-438-52-3914)
FEATURES
    source
        1. 5120
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /chromosome="2"
            /clone="h307423"
            /tissue_type="brain"
            /clone_lib="pBluescriptII SK plus"
        2. 5120
            /gene="KIAA1083"
            /gene="KIAA1083"
            /codon_start=1
            /product="KIAA1083 protein"
            /protein_id="BAA83035.1"
            /db_xref="GI:5689503"
            /translation="MNSPGGRKKKGGSGASNPVPRPPPCJAPAPPAGPAPPES
            PHKRNLYFYSYPLFVGFALLRLVAFHLGLFVWLCQRFSRALMAKRSSGAPAPASA
            SAPAPVGGGEAEVRVVFHKQAFYISIALRIDEKAGQKEQAVENYKNGISELEKGI
            AVIVTQGEQGERARRLQAKMYNLVMAKORLQLEKMPVLPEKSKQTDVNDSTN
            ACNGHLSQESGAVPKRKOPLTHTSNSLPRSKTMKTSAGSGHHRAPSYSGLSMVS
            GVKGSQSPAPTHKGIPTKNTNKPSTPTTATRKKKOLKFNVDENLANLKNEIVD
            NGTAVKFDIAGQDLAKQALQEIIVLPSLRPELFTGLRAPARGLLFPGPGNGKTMIA
            KAVAAESNATFENSAASLTSKYVGEKLVRLFAVARELQFSIFIDEVLSLQCR
            REGHDAARLLKTEFLIEFDGVQSGADDRVLVVGATNRPOELDEAVLRRLKRVYVSL
            PNEETRLLLKNTLCKGSLTQKELACLARVTDGVSGSLTALAKDAAGPIRELKP
            EQVKNMSASEMRNRLSDFTESLKKIKRSVSPQLEAYIRWKKDFGDTTV"
            polyA_signal 3227..3232
            /gene="Spast"
BASE COUNT 1532 a 943 c 1077 g 1568 t
ORIGIN
Query Match 40.3%; Score 403; DB 9; Length 5120;
Best Local Similarity 100.0%; Pred. No. 5.5e-67;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 599 ACATGAGAAATATTCGATTATCTGACTTCATGAACTCTTGAAAAAATAAAGCGGCG 658
```

Db 1852 AGATCAGAAATATTTCGATTATCTGACTTCACCTGAATCCTTGAAGAAAATAAAACGCAGCG 1911

QY 659 TCAGCCCTCAAACTTTAGAGCGGTACATACGTTGGAAACAAGGACTTTGGAGATACCACTG 718

Db 1912 TCAGCCCTCAAACTTTAGAGCGGTACATACGTTGGAAACAAGGACTTTGGAGATACCACTG 1971

QY 719 TTTAAGGAAATACCTTTGTAAAGCTGCAGAACATTTTACTTAAAGAGGAAACACAAAGAT 778

Db 1972 TTTAAGGAAATACCTTTGTAAAGCTGCAGAACATTTTACTTAAAGAGGAAACACAAAGAT 2031

QY 779 CTTCAAATGAACGTTCATCGGCTACAGAAACAGCCTTAAGTTTACAGGACTTTTACAGGCTTT 838

Db 2032 CTTCAAATGAACGTTCATCGGCTACAGAAACAGCCTTAAGTTTACAGGACTTTTACAGGCTTT 2091

QY 839 ACATATTTGTGCACCAAACTTTGAAGATGAACACAGAAACACACTTAAACAAAATATACAA 898

Db 2092 ACATATTTGTGCACCAAACTTTGAAGATGAACACAGAAACACACTTAAACAAAATATACAA 2151

QY 899 TGCAAATGTAAATTTTGTGTTTAAAGGCTTCGCTTGATGTCACAGTTATCCCAATGG 958

Db 2152 TGCAAATGTAAATTTTGTGTTTAAAGGCTTCGCTTGATGTCACAGTTATCCCAATGG 2211

QY 959 ACCTAAGTTAGACCAACAACAACTGATTCGTGCTCTCTT 1001

Db 2212 ACCTAAGTTAGACCAACAACAACTGATTCGTGCTCTCTT 2254

RESULT 8

AC093899/c

LOCUS

DEFINITION Homo sapiens BAC clone RP11-724016 from 2, complete sequence.

ACCESSION AC093899 AC068884

VERSION AC093899.3 GI:18497265

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 172816)

Su1ston,J.E. and Waterston,R.

Toward a complete human genome sequence

Genome Res. 8 (11), 1097-1108 (1998).

99063792

9847074

PUBMED

2 (bases 1 to 172816)

Pearman,C., Haakenson,W. and Boyer,E.

The sequence of Homo sapiens BAC clone RP11-724016

Unpublished (2001)

3 (bases 1 to 172816)

Waterston,R.H.

Direct Submission

Submitted (10-SEP-2002) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA

4 (bases 1 to 172816)

Waterston,R.H.

Direct Submission

Submitted (05-FEB-2002) Genome Sequencing Center, Washington University School of Medicine, 444 Forest Park Parkway, St. Louis, MO 63108, USA

5 (bases 1 to 172816)

Waterston,R.

Direct Submission

Submitted (12-JUN-2002) Department of Genetics, Washington University, 444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Feb 5, 2002 this sequence version replaced gi:15625013.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: sapiens@watson.wustl.edu

----- Summary Statistics

Center project name: H_NH0724016

Drafting Center: WGBR

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tatem,M., Catanese,J.C. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries: Genomics 5:11-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-724016; actual end is at base position 172816 of RP11-724016.

Unresolved simple sequence repeat from base position 22700 to 25900.

The sequence of AC068884 has been incorporated into AC093899.

FEATURES

Source

1..172816

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="2"

/map="2"

/clone="RP11-724016"

/clone_lib="RPC2-11"

19..1140

repeat_region

/rpt_family="L1"

1141..1437

repeat_region

/rpt_family="Alu"

1438..2657

repeat_region

/rpt_family="L1"

2658..2729

repeat_region

/rpt_family="L1"

2730..3320

repeat_region

/rpt_family="L1"

3321..3642

repeat_region

/rpt_family="Alu"

3643..4184

repeat_region

/rpt_family="L1"

4185..4996

repeat_region

/rpt_family="ERV1"

5011..5263

repeat_region

/rpt_family="L1"

5285..5907

repeat_region

/rpt_family="CR1"

6895..7078

repeat_region

/rpt_family="L1"

7103..7421

repeat_region


```
repeat_region /rpt_family="Alu"
7807. .7952
repeat_region /rpt_family="L1"
7981. .8289
repeat_region /rpt_family="Alu"
8360. .8551
repeat_region /rpt_family="L1"
8622. .8718
repeat_region /rpt_family="MIR"
8784. .8825
repeat_region /rpt_family="(TTTTC)n"
8840. .9134
repeat_region /rpt_family="Alu"
9136. .9174
repeat_region /rpt_family="Alu"
9861. .9881
repeat_region /rpt_family="(A)n"
9882. .10166
repeat_region /rpt_family="Alu"
16502. .16556
repeat_region /rpt_family="L2"
16886. .17165
repeat_region /rpt_family="L1"
17179. .17482
repeat_region /rpt_family="Alu"
18025. .18119
repeat_region /rpt_family="MIR"
18177. .18471
repeat_region /rpt_family="Alu"
19312. .19441
repeat_region /rpt_family="MER1_type"
19731. .20040
repeat_region /rpt_family="MaLR"
20403. .20472
repeat_region /rpt_family="L2"
21286. .21344
repeat_region /rpt_family="(CA)n"
22724. .23977
repeat_region /rpt_family="(TA)n"
24005. .24171
repeat_region /rpt_family="(TA)n"
24196. .25007
repeat_region /rpt_family="(TA)n"
25030. .25225
repeat_region /rpt_family="(TA)n"
25223. .25398
repeat_region /rpt_family="(TA)n"
25399. .25455
repeat_region /rpt_family="AT-rich"
25456. .25919
repeat_region /rpt_family="(TA)n"
25920. .26206
repeat_region /rpt_family="Alu"
26893. .26917
repeat_region /rpt_family="AT-rich"
27194. .27497
repeat_region /rpt_family="L1"
27498. .27814
repeat_region /rpt_family="MER2_type"
27815. .28151
repeat_region /rpt_family="L1"
28152. .28278
repeat_region /rpt_family="Alu"
28279. .28485
repeat_region /rpt_family="L1"
28795. .28829
repeat_region /rpt_family="AT-rich"
28986. .29100
repeat_region /rpt_family="L2"
29237. .29349
repeat_region /rpt_family="L1"
29350. .29484
repeat_region /rpt_family="Alu"
```

```
repeat_region 29485. .29589
/rpt_family="L1"
repeat_region 29811. .29874
/rpt_family="L2"
repeat_region 29875. .30284
/rpt_family="L1"
repeat_region 30320. .30522
/rpt_family="L1"
repeat_region 30523. .30847
/rpt_family="Alu"
repeat_region 30846. .30862

Query Match 20.0%; Score 200.6; DB 9; Length 172816;
Best Local Similarity 60.8%; Pred. No. 1.2e-28;
Matches 398; Conservative 0; Mismatches 249; Indels 9; Gaps 4.

QY 5 TAAAAATGCATACCTTTAGGCTGGGCAAAAGTGGTCACAGTCTGTGTAATCCACAGCACTTTGGG 64
DB ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 65 AGGCCAAGGTGGGGGATCACCTGAGGTTAGGAGTTAGGAGTTCAGGACCAACCTGGCCATCATGGC 124
DB ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 26163 AGGCTGAGGTGGGCAGATCACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCATATAGT 26104
QY 125 GAAACCCCTGTCTCTATATAAAATACAAAAAATTAGTGGGCATGGAGGCATGTGCGTGTA 184
DB ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 26103 GAAACTTCGTCTCTACTAAAAATACAAAAAATTAGCAAGCATGGTGGTGTGTGCGCTGTAA 26044
QY 185 TCCCAGCTACTCGGGAGGCTGAGACAGAGAATGCTTTGAACCTTGGAGGGGGAGGTTGC 244
DB ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 26043 TCCCAGCTACTCGGGAGGCTGAGACAGAGAATGCTTTGAACCTTGGAGGGGGAGGTTGC 25964
QY 245 ATATCTGAGTGGTGAATTTGTGATCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 304
DB ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 25983 AT---TGAGTGGAGAGAGTGGCCACTGGCCCTCCAGTCTGGGTAAACAGAGACATCCCATC 25927
QY 305 TATAAATGATTTGTGTTGTTTTTTTATATTGGAAAAATATTATGCTTTCAAATGTTAATAAC 364
DB ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 25926 TAAAAAATATATATATATAAATATATATAAATATATATAAATATATAAATATATAAATATA 25867
QY 345 CTATGAACATAACACAGTAATAAATATATTATAGTATAGCATTTTATTAAGGTTTCTTGTG 424
DB ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 25866 ATATATAATATATATATATATATATATATATATATATATATATATATATATATATATATA 25807
QY 425 TAGCAGATCAACATAGAAATATATTTAAATGGGTGACATAATTTCTA---AGAAATACA 481
DB ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 25806 TA-AATATATATATAAATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAACA 25748
QY 482 TACACGTATATTTTTTATAACAT-TAAGAAACAGCAGCATCATTTTAAATCCATCATTT 540
DB ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 25747 TATTAATATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 25688
QY 541 TCGTTAACCAACCATATACCTGTGTGATCATTTGTATTTGTCATGTGCTTTTAAAAATCTAG 600
DB ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 25687 AATTTATATATATTACATATATATTATATATAAATATATATATATATATATATATATATATAT 25628
QY 601 ATGAGAAATATTCGATTTCTGACTTCATGAATCCCTTGAAAAAATAAACCCTA 655
DB ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 25627 ATACAAATATATAAAATTTATATATATATATATATTTACATATATAAATAAATAAATGTA 25573
```

```
RESULT 9
AC130403
LOCUS
DEFINITION Homo sapiens 12 BAC RP11-674N14 (Roswell Park Cancer Institute
Human BAC Library) complete sequence.
ACCESSION AC130403
VERSION AC130403.3 GI:24431574
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
```



```
repeat_region      /rpt_family="LIPB1"
6477..6520
/rpt_family="TG1A"
6507..6688
/rpt_family="TA1A"
complement(6511..6702)
/rpt_family="LIPB1"
6703..6762
/rpt_family="TA1A"
complement(6775..7080)
/rpt_family="LIPB1"
7083..8073
/rpt_family="LIPB"
8079..8224
/rpt_family="LIP5"
complement(8344..8577)
/rpt_family="ML1K"
8666..8708
/rpt_family="ATG1A"
8779..8877
/rpt_family="L2"
complement(9302..9463)
/rpt_family="FRAM"
complement(9468..9507)
/rpt_family="MER116"
9558..9839
/rpt_family="AluSp"
9840..9883
/rpt_family="(TAA)n"
complement(10135..10518)
/rpt_family="MLT1B"
10795..10822
/rpt_family="AT_rich"
11068..11241
/rpt_family="LIP5"
11272..11509
/rpt_family="LIP12"
11641..11789
/rpt_family="LIP3"
12052..12332
/rpt_family="AluSg"
12333..12383
/rpt_family="(TAA)n"
14051..14943

Query Match      20.6%  Score 199.8;  DB 9;  Length 195432;
Best Local Similarity 88.9%;  Pred. No. 1,70-28;
Matches 216;  Conservative 0;  Mismatches 27;  Indels 0;  Gaps 0;

QY  3  TATAAAATGCATCTTTAGGCTGGGCAAGTGGCTCAGCTGTGTATATCCAGCACTTIG 62
DB  44699  TAGAAAACCTCCATATCTAGGCTGGGCAAGTGGCTCAGCTGTGTATATCCAGCACTTIG 44758
QY  63  GGAGGCAAGGTGGGCGGATCAGCTGAGGTTAGGAGTTCAGGACCAACCTGGGCACTATG 122
DB  44759  GGAGGCTGAGCGGGTGGATCAGCTGAGGTTAGGAGTTCAGGACCAACCTGGGCACTATG 44818
QY  123  GCGAAACCCCTGCTCTACTATAAATACAAAATAGCTGGGATGGAGGCAATGGCCCTGT 182
DB  44819  GTGAACCCCTGCTCTACTATAAATACAAAATAGCTGGGATGGAGGCAATGGCCCTGT 44878
QY  183  AATCCAGCTACTCGGAGGCTGAGACAGAGAAATTCCTTGAACCTTGGAGGGGAGGTT 242
DB  44879  AATCCAGCTACTCGGAGGCTGAGGAGGAGAAATTCCTTGAACCCAGGAGGGGAGGTT 44938
QY  243  GCA 245
DB  44939  GCA 44941

RESULT 10
AC068752/c
LOCUS      AC068752      209354 bp      DNA      linear      HTG 12-AUG-2000
```

```
DEFINITION      Homo sapiens chromosome 12 clone RP11-674N14, WORKING DRAFT
SEQUENCE, 37 unordered pieces.
ACCESSION      AC068752
VERSION      AC068752.2  GI:9798052
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 209354)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 209354)
Waterston,R.H.
Direct Submission
Submitted (28-MAY-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Aug 12, 2000 this sequence version replaced gi:7717179.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0674N14
----- Summary Statistics -----
Sequencing vector: M13, 100%
Sequencing vector: plasmid; 0%
Chemistry: Dye-primer ET; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 189469 bases at least Q40
Consensus quality: 194965 bases at least Q30
Consensus quality: 197585 bases at least Q20
Insert size: 205000; agarose-fp
Insert size: 205754; sum-of-contigs
Quality coverage: 3.54 in Q20 bases; agarose-fp
Quality coverage: 3.63 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 37 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1492: contig of 1492 bp in length
* 1493: gap of unknown length
* 1593: contig of 1876 bp in length
* 3469: gap of unknown length
* 3569: contig of 2453 bp in length
* 6022: gap of unknown length
* 6122: contig of 2229 bp in length
* 8350: gap of unknown length
* 8351: gap of unknown length
* 8451: contig of 1284 bp in length
* 9735: gap of unknown length
* 9835: contig of 2760 bp in length
* 12595: gap of unknown length
* 12695: contig of 2020 bp in length
* 14715: gap of unknown length
* 14815: contig of 2580 bp in length
* 17395: gap of unknown length
* 17495: contig of 3382 bp in length
* 20877: gap of unknown length
* 20977: contig of 3377 bp in length
* 24354: gap of unknown length
* 24454: contig of 3462 bp in length
* 27916: gap of unknown length
* 28016: contig of 3598 bp in length
* 31614: gap of unknown length
```

```
REFERENCE
AUTHORS
JOURNAL
REFERENCE
AUTHORS
JOURNAL
COMMENT
```



```

QY      243 GCA 245
Db      201974 GCA 201872

AC002549      163027 bp      DNA      linear      PRI 27-MAR-1998
LOCUS      Homo sapiens Xp22 BAC SS-377014 (Genome Systems Human BAC library)
DEFINITION      complete sequence.
ACCESSION      AC002549
VERSION      AC002549.1 GI:2992475
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 163027)
AUTHORS      Muzny,D., Arenson,A.D., Brundage,E., Carvelli,K., Chen,E., Chen,C.,
Di,W., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Ganesh,R.,
Garcia,C., Goodman,M., Gorrell,J.H., Haywood,M., Jackson,J.,
Jin,S., Karpal,R., Karpathy,S., Leal,B., Li,Y., Liu,W., Logan,C.,
Lu,C., Ly,T., Martinez,C., Oswal,G., Perez,J., Rashid,N.D.,
Rowland,K., Savage,L., Scherer,S.S., Shen,H., Timms,K.M., Todd,J.,
Vo,Q., Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.
Direct Submission
Unpublished
REFERENCE      2 (bases 1 to 163027)
AUTHORS      Chiu,X.W.
Direct Submission
Submitted (18-SEP-1997) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE      3 (bases 1 to 163027)
AUTHORS      Worley,K.C.
Direct Submission
Submitted (27-MAR-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT      On Mar 27, 1998 this sequence version replaced gi:2960513.
Sequencing is completed to a minimum standard of double strand
coverage with a minimum of 2 clones and 2 reads with no ambiguities
or 2 chemistries with a minimum of 2 clones and 3 reads with no
ambiguities. If the sequence quality does not meet this standard,
it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by
Adrian S-It.

Sequence similarities were identified using Powerblast by Jinghui
Zhang.

Exon/intron boundaries of identified genes were chosen if there
were canonical splice junctions that maintained sequence continuity
across the splice junctions.
Location/Qualifiers
..163027
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xp22"
/clone="SS-377014"
/clone_lib="Genome Systems Human BAC library"
1..171
/note="Overlap bases 123033-123203 in AC003036"
525..823
/rpt_family="AluSx"
repeat_region
891..1742
/rpt_family="L1M4"
repeat_region
1776..1970
/rpt_family="L1PA13"
repeat_region
2071..2159
repeat_region

misc_feature
/note="Region: W0444C za42333.r1 Soares fetal liver spleen
INFLS Homo sapiens CDNA clone 295349 5'"
repeat_region
complement(20265..20375)
/rpt_family="L1XAI"
20439..20574
/rpt_family="MER33"
repeat_region
complement(20575..20619)
/rpt_family="AT_rich"
complement(21136..21502)
/misc_feature
/note="Region: R98637 yc69h08.s1 Soares fetal liver spleen
INFLS Homo sapiens CDNA clone 201087 3'"
21353..21502
/standard_name="WI-12677, Chr. 1, Homo sapiens"
/db_xref="dbSTS:21718"
23674..23770
/rpt_family="L1CA1"
repeat_region
23689..23816
/standard_name="AFMa067ydi, Chr. X, Homo sapiens"
/db_xref="dbSTS:29765"

```

```

/rpt_family="(TA)n"
2232..2461
/rpt_family="AluJb"
2480..2535
/rpt_family="POLY_A"
complement(2541..2741)
/rpt_family="MER2"
3707..3915
/rpt_family="MER58A"
4261..4562
/rpt_family="AluSx"
4704..4916
/rpt_family="MER20"
complement(7185..7450)
/rpt_family="AluSg"
7669..7708
/rpt_family="NIR"
8605..8646
/rpt_family="AT_rich"
complement(9401..9702)
/rpt_family="AluJb"
9764..9950
/rpt_family="MER58B"
10044..10306
/rpt_family="L1MB6"
10309..10438
/rpt_family="AluSx"
10440..10466
/rpt_family="(CAAA)n"
10467..10651
/rpt_family="AluSx"
10669..10941
/rpt_family="L1MB3"
11888..12010
/rpt_family="L2"
12400..12714
/rpt_family="L2"
complement(12411..14713)
/rpt_family="AluSx"
15004..15225
/rpt_family="MER20"
17071..17147
/rpt_family="L2"
18312..18457
/rpt_family="NIR"
complement(18528..18821)
/rpt_family="AluJb"
complement(19102..19403)
/rpt_family="AluSx"
complement(19406..19587)
/rpt_family="MER33"
complement(19666..19752)
/rpt_family="MER3"
20048..20304
/note="Region: W0444C za42333.r1 Soares fetal liver spleen
INFLS Homo sapiens CDNA clone 295349 5'"
complement(20265..20375)
/rpt_family="L1XAI"
20439..20574
/rpt_family="MER33"
complement(20575..20619)
/rpt_family="AT_rich"
complement(21136..21502)
/standard_name="WI-12677, Chr. 1, Homo sapiens"
/db_xref="dbSTS:21718"
23674..23770
/rpt_family="L1CA1"
repeat_region
23689..23816
/standard_name="AFMa067ydi, Chr. X, Homo sapiens"
/db_xref="dbSTS:29765"

```

```

repeat_region 24776..25005
/rpt_family="L1MD3"
repeat_region 25007..25295
/rpt_family="AluJo"
repeat_region 25303..25344
/rpt_family="POLY_A"
repeat_region 25918..26006
/rpt_family="L2"
repeat_region complement(25409..26625)
/rpt_family="AluCo"
repeat_region 26983..27190
/rpt_family="MLTIF"
repeat_region 29007..29429
/rpt_family="MLTIC"
repeat_region 29637..29722
/standard_name="DXS1313"
repeat_region 30740..30771
/rpt_family="POLY_A"
repeat_region complement(31044..31336)
/rpt_family="AluSx"
repeat_region 32605..32971
/rpt_family="MLT1A1"
repeat_region 33703..33863
/rpt_family="MLTIC"
repeat_region 33864..34165
/rpt_family="AluSx"
repeat_region 34166..34197
/rpt_family="(CAAAA)n"
repeat_region 34210..34481
/rpt_family="MLTIC"
repeat_region complement(34829..35177)
/rpt_family="MLTIB"
repeat_region 36433..36519
/rpt_family="MER20"
repeat_region complement(36567..36706)
/rpt_family="MER20"
repeat_region complement(39685..39987)
/rpt_family="AluSq"
repeat_region 40270..40647
/rpt_family="L1MB4"
repeat_region 41624..41651
/rpt_family="Alu_rich"
repeat_region complement(42747..43322)
/rpt_family="L1MD3"
repeat_region 43727..43910
/rpt_family="MLTIC"
repeat_region 44205..44499
/rpt_family="AluSx"
repeat_region 44763..44780
/rpt_family="L1MB7"
repeat_region 44782..45095
/rpt_family="MER2"
repeat_region 45127..45474
/rpt_family="L1MB7"
repeat_region complement(46188..46488)
/rpt_family="AluJo"
repeat_region complement(46489..47116)
/rpt_family="L1MB2"
repeat_region complement(47166..47209)
/rpt_family="Mir"
repeat_region 47211..47553
/rpt_family="MLTIC"

```

JOURNAL Submitted (17-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 163662)

AUTHORS Worley, K.C.

TITLE Direct Submission

JOURNAL Submitted (13-SEP-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT On Sep 13, 2002 this sequence version replaced gi:20376557.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Drafting Center Code: BCM
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: HDMC
Center clone name: RP11-111F17
----- Summary Statistics -----
Sequencing vector: Plasmid;
Sequencing vector: M13;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 161410 bases at least Q40
Consensus quality: 162039 bases at least Q30
Consensus quality: 162473 bases at least Q20
Estimated insert size: 164107; sum-of-contigs estimation
Quality coverage: 10x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence, as soon as it is available and the accession number will be preserved.

1 3227: contig of 3227 bp in length
3228 3227: gap of unknown length
3229 3227: gap of 9326 bp in length
12653: contig of 9326 bp in length
12654 12653: gap of unknown length
12754 33426: contig of 20673 bp in length
33427 33526: gap of unknown length
33527 60992: contig of 27465 bp in length
60992 61091: gap of unknown length
61092 163662: contig of 102571 bp in length.

Location/Qualifiers
1. 163662
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/clone="RP11-111F17"

BASE COUNT 43922 a 36511 c 35209 g 44615 t 405 others

ORIGIN

Query Match
Best Local Similarity 88.8%; Score 197.8; DB 2; Length 163662;
Matches 214; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 5 TAAAAATGCATCTTTAGGCTGGGCAAGTGGCTACGCTGTGTAAATCCAGCACATTTGGG 64
|||||
Dd TAAAAATGGTTAAAAATAGGCGGGCACACTGGCTCAGGCTGTAAATCCAGCACATTCGGG 155497
|||||

QY 65 AGGCCAAGTGGCGGATCACCTGAGGTTAGGAGTTACAGACCAACCTGGCCATCATGGC 124
|||||
Dd AGGCCAAGTGGGTGGATCACCTGAGGTTACAGAGTTTACAGACCAAGCTGGCCAAATGGT 155437
|||||

QY 125 GAACCCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCTGTAA 184
|||||

Dd 155436 GAAACCCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGTGGGTGGCTGTGCTGTAA 155377
|||||

QY 185 TCCAGCTACTCGGAGGCTGAGACAGAGAAATTCCTTGAACCTTGGAGGCGAGGTTCG 244
|||||

Dd 155376 TCCAGCTACTCAGGAGGCTGAGGAGGAGAAATTCCTTGAACCTTGGAGGCGAGGTTCG 155317
|||||

QY 245 A 245
|||||

Dd 155316 A 155316

RESULT 13
ACC68676/c
LOCUS
DEFINITION Homo sapiens clone RP11-768N8, WORKING DRAFT SEQUENCE. 14 unordered pieces.
ACCESSION ACC68676
VERSION ACC68676.3 GI:11875293
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 191866)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Bedalov, A., Boguslavskiy, L., Bouckegalter, B., Brown, A., Burkett, S., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Sage, D., Galagar, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, S., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McSheeters, R., Meldrum, J., Merens, L., Mihova, T., Miranda, C., Miengo, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, S., Peterson, K., Pierre, N., Pisani, C., Polara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, C., Vassiliou, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, E., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-MAY-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Dec 16, 2002 this sequence version replaced gi:10440660.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center -----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L10325
Center clone name: 768 N 8
----- Summary Statistics -----
Sequencing vector: Plasmid; n/a; 98% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 186020 bases at least Q40
Consensus quality: 188219 bases at least Q30
Consensus quality: 189180 bases at least Q20
Insert size: 188000; agarose-fp

Insert size: 193566; sum-of-contigs
 Quality coverage: 9.2 in Q20 bases; agarose-
 Quality coverage: 9.1 in Q20 ba.
 NOTE: This is a 'working draft' sequence. It currently
 consists of 14 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.

1 527: contig of 527 bp in length
 528 627: gap of 100 bp
 628 1287: contig of 660 bp in length
 1298 1387: gap of 100 bp
 1388 1833: contig of 446 bp in length
 1834 1934: gap of 100 bp
 1934 2723: contig of 790 bp in length
 2724 2823: gap of 100 bp
 2824 4471: contig of 1648 bp in length
 4472 4571: gap of 100 bp
 4572 8184: contig of 3613 bp in length
 8185 8285: gap of 100 bp
 8285 18881: contig of 10597 bp in length
 18882 18981: gap of 100 bp
 18982 33829: contig of 14848 bp in length
 33830 33929: gap of 100 bp
 33930 51797: contig of 17868 bp in length
 51798 51897: gap of 100 bp
 51898 67219: contig of 15322 bp in length
 67220 67319: gap of 100 bp
 67320 88231: contig of 20912 bp in length
 88232 88332: gap of 100 bp
 88332 132689: contig of 44356 bp in length
 132690 132789: gap of 100 bp
 132790 170094: contig of 37305 bp in length
 170095 170194: gap of 100 bp
 170195 191866: contig of 21672 bp in length.

FEATURES

source
 1. 191866
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /clone_id="RP11-768N8"
 /clone_lib="RPC1-11 Human Male BAC"
 misc_feature
 1. 527
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 misc_feature
 628. 1287
 /note="assembly_fragment"
 misc_feature
 1388. 1833
 /note="assembly_fragment"
 misc_feature
 1934. 2723
 /note="assembly_fragment"
 misc_feature
 2824. 4471
 /note="assembly_fragment"
 misc_feature
 4572. 8184
 /note="assembly_fragment"
 misc_feature
 8285. 18881
 /note="assembly_fragment"
 misc_feature
 18982. 33829
 /note="assembly_fragment"
 misc_feature
 33930. 51797
 /note="assembly_fragment"
 misc_feature
 51898. 67219
 /note="assembly_fragment"
 misc_feature
 67320. 88231
 /note="assembly_fragment"
 misc_feature
 88332. 132689
 /note="assembly_fragment"
 misc_feature
 132790. 170094
 /note="assembly_fragment"

misc_feature 170195..191866
 /note="assembly_fragment"
 clone_end:R7
 vector_side:right

BASE COUNT 52298 a 42689 c 42735 g 52841 t 1303 others
 ORIGIN

Query Match 19.8%; Score 197.8; DB 2; Length 191866;
 Best Local Similarity 88.8%; Pred. No. 4e-28;
 Matches 214; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 5 TAAAAATGCATACITTAGCTGGGCAAAAGTGGCTCACGTCCTGTAAATCCAGCACTTTGGG 64
 |||||
 Db 33796 TAAAAATGGTTAAATATAGGCGGGCACACTGGCTCACGCTGTAAATCCAGCACTTGGG 33737
 QY 65 AGGCCAAGSTGGGCGGATCACCTGAGTTAGGAGTTCAGGACCAACCTGGCCATCATGGC 124
 |||||
 Db 33736 AGGCCAAGSTGGTGGATCACCTGAGGTCAGGAGTTTGAACCCAGCTGGCCACATGGT 33677
 QY 125 GAAACCGCTGCTCTACTAAAAATACAAAAATTAGTGGGCTAGGAGGCAATGTGCTGTAA 184
 |||||
 Db 33676 GAAACCGCTGCTCTACTAAAAATACAAAAATTAGTGGGCTAGGAGGCTGTAA 33617
 QY 185 TCCAGCTACTCGGGAGGCTGAGACAGAGAATTCCTTGAACCTTGGAGGGGAGGTTGC 244
 |||||
 Db 33616 TCCAGCTACTCAGGAGGCTGAGGAGGAGAGATTGCTTGAACCTTGGAGGAGGTTGC 33557
 QY 245 A 245
 Db 33556 A 33556

RESULT 14
 AC099804

LOCUS Homo sapiens chromosome 17, clone CTC-2357A8, complete sequence.
 DEFINITION
 AC099804

AC099804.3 GI:22532758

KEYWORDS

SOURCE

Homo sapiens (human).

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS Birren, B., Nusbaum, C. and Lander, E.

TITLE Homo sapiens chromosome 17, clone CTC-2357A8

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 19322)

AUTHORS

Birren, B., Linton, J., Nusbaum, C., Lander, E., Allen, A., Allen, N.,
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
 Brown, A., Camarata, C., Campolanci, A., Chang, J., Chazaro, B.,
 Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
 Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Faru, S.,
 Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
 Jones, C., Kamat, A., Karatas, A., Kellis, C., LaRoque, K.,
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
 MacLean, C., Macdonald, P., Mavor, J., Marquis, N., Matthews, C.,
 McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrum, J.,
 Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
 Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
 Oliver, J., Peterson, K., Phurkhang, P., Pierre, N., Pollara, V.,
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Strauss, N., Subramanian, A., Talamas, C., Tesfaye, S., Theodore, J.,
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

TITLE

JOURNAL

Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA


```

REFERENCE
AUTHORS
3 (bases 1 to 119922)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouckhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepe,Y., Collymore,A.,
Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,H.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,G., Matthews,C.,
McCarthy,M., Meldrim,C., Meneus,L., Mihova,T., Mierga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talaras,C.,
Tesfaye,S., Theodore,J., Topham,K., Travers,N., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
4 (bases 1 to 119922)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Bouckhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepe,Y., Collymore,A.,
Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,H.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., MacLean,C., Macdonald,P., Major,G., Matthews,C.,
McCarthy,M., Meldrim,C., Meneus,L., Mihova,T., Mierga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talaras,C.,
Tesfaye,S., Theodore,J., Topham,K., Travers,N., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (28-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 28, 2002 this sequence version replaced gi:22391567.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RV/RepeatMasker.html
----- Genome Center
----- Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 210689
Center clone name: 2357_A_8
-----
FEATURES
source
1. 119922
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="CTD-2357A8"
/clone_lib="CITD1 Human BAC"
280..573
/rpt_family="AluSx"
682..784
/rpt_family="L2"
790..1531
/rpt_family="L1PA13"
2246..2386
repeat_region
repeat_region
repeat_region
repeat_region
/rpt_family="MLT2B2"
2752..2953
/rpt_family="AluY"
3116..3248
/rpt_family="MIR3"
complement(3357..3446)
/rpt_family="AluSP/q"
complement(3447..3532)
/rpt_family="AluSx"
complement(5336..5513)
/rpt_family="AluSg/x"
complement(5738..6036)
/rpt_family="AluJb"
6792..7153
/rpt_family="THE1B"
7223..7425
/rpt_family="MSTA"
7426..7721
/rpt_family="AluSg"
7722..7905
/rpt_family="MSTA"
complement(7906..8219)
/rpt_family="AluSg"
8220..8289
/rpt_family="MSTA"
8302..8483
/rpt_family="MER77"
9281..9385
/rpt_family="HERV16"
complement(9386..9520)
/rpt_family="FLAM_C"
9521..9655
/rpt_family="HERV16"
complement(10817..11516)
/rpt_family="MER50"
12317..12337
/rpt_family="AT-rich"
12488..13151
/rpt_family="L1MB6"
13152..13456
/rpt_family="AluSx"
13457..13674
/rpt_family="L1MB6"
complement(13691..13832)
/rpt_family="AluJb"
13875..14322
/rpt_family="L1MB5"
16058..16097
/rpt_family="L1M4"
16992..17013
/rpt_family="(Alu)"
17086..17462
/rpt_family="L1MB5"
17464..17668
/rpt_family="MER2"
17705..17806
/rpt_family="AluS"
17819..17895
/rpt_family="MER2"
complement(17916..18220)
/rpt_family="AluSx"
18221..18367
/rpt_family="L1MB5"
18811..18876
/rpt_family="L1MB3"
18883..18953
/rpt_family="L1MB8"
complement(18955..19262)
/rpt_family="AluY"
18993..19040
-note="PCR product sequence only"
19346..19752
/rpt_family="MLT1C"

```

unsure 19361 /note="Probably G; possibly T"
unsure 19412..19417 /note="<30 qual SNGL region"
unsure 19459..19530 /note="single clone coverage"
repeat_region 20040..20337 /rpt_family="AluSp"
unsure 20341..20345 /note="<30 qual SNGL region"
unsure 20425..20443 /note="<30 qual SNGL region"
repeat_region 20450..20745 /rpt_family="AluSc"
unsure 20495..20503

Query Match 19.7%; Score 196.8; DB 9; Length 119922;
Best Local Similarity 88.8%; Pred. No. 1.4e-28;
Matches 213; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 6 AAAATGTCATCTTAGGCTGGGCAAGTGGCTGACGTCTGTGTAATCCAGCACTTTGGGA 65
Db 7410 AAAAGTGTGTAGCACTGGCTGGGACAGTGGCTGACGCTGTATATCCAGCACTTTGGGA 7469
QY 66 GGCCAGGTGGGCGATCACCTGAGGTAGGATTCAGGACCAACCTGGGCAATCATGGCG 125
Db 7470 GGCCAGGTGGGCGATCACCTGAGGTAGGATTCAGGACCAACCTGGGCAATCATGGCG 7529
QY 126 AAACCTGTCTCTACTAAATAACAAAATTAGCTGGGATGGAGGCAATGCTGTAAAT 185
Db 7530 AAACCTGTCTCTCTACTAAATAACAAAATTAGCTGGGATGGAGGCAATGCTGTAAAT 7589
QY 186 CCCAGTACTCGGAGGCTGAGACAGAGGATTCCTGACCTGGAGGAGGAGGTTGGA 245
Db 7590 CCCAGTACTCGGAGGCTGAGACAGAGGATTCCTGACCTGGAGGAGGAGGTTGGA 7649

RESULT 15
HSJ738A13/c HSJ738A13 46775 bp DNA linear PRI 01-FEB-2003
LOCUS Human DNA sequence from clone RP4-738A13 on chromosome
DEFINITION Xq21.31-22.1, complete sequence.
ACCESSION AL109801
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 46775)
Whitehead, S.
Direct Submission
Submitted (31-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 1, 2003 this sequence version replaced gi:6624938.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EMBL, EMBL; SWISSPROT, TrEMBL; TrEMBL; WP, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
Chromosome X, constructed by the Sanger Centre Chromosome X Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
----- Genome Center
Center: Wellcome Trust Sanger Institute

Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at
least one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
RP4-738A13 is from the library RPCI-4 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2.

FEATURES
Location/Qualifiers
1..46775
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="X"
/map="q21.31-22.1"
/clone="RP4-738A13"
/C-one_11b="RPCI-4"
BASE COUNT 13491 a 9466 c 9528 g 14290 t
ORIGIN

Query Match 19.6%; Score 195.8; DB 9; Length 46775;
Best Local Similarity 88.7%; Pred. No. 1.1e-27;
Matches 212; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 7 AAATGTCATCTTAGGCTGGGCAAGTGGCTGACGTCTGTAAATCCAGCACTTTGGGAG 66
Db 39786 AAAAGTATATAATTAGTCCGGGCAAGTGGCTGACCTGTAAATCCAGCACTTTGGGAG 39723
QY 67 GCGAAGGTGGGCGGATCACCTGAGGTAGGATTCAGGACCAACCTGGGCAATCATGGCGA 126
Db 39726 GCGAAGGTGGGCGGATCACCTGAGGTAGGATTCAGGACCAACCTGGGCAATCATGGCGA 39667
QY 127 AACCTGTCTCTACTAAATAACAAAATTAGCTGGGATGGAGGCAATGCTGTAAATC 186
Db 39666 AACCTGTCTCTACTAAATAACAAAATTAGCTGGGATGGAGGCAATGCTGTAAATC 39607
QY 187 CCAGCTACTCGGAGGCTGAGACAGAGGATTCCTGACCTGGAGGAGGAGTTGGA 245
Db 39606 CCAGCTACTCGGAGGCTGAGACAGAGGATTCCTGACCTGGAGGAGGAGTTGGA 39548

Search completed: October 24, 2003, 21:06:17
Job time : 2770.61 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:39:20 : Search time 252.249 seconds
(without alignments)
10712.169 Million cell updates/sec

Title: US-09-830-902-1_COPY_100000_101000

Perfect score: 1001

Sequence: 1 aatataaaatgcatactttt.....acotgatctggtctctttt 1001

Scoring table: IDENTITY_NJC

Gapop 10.0, Gapext 1.0

Searched: 2512756 seqs, 1343719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*
- 25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2004.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100.1	100.0	10000	22	AAF84800
2	40.3	40.3	3263	22	AAF84801
3	40.3	40.3	4152	24	ABQ54772
4	191.4	19.1	3143	24	AAS17746
5	191.4	19.1	30573	22	AALC4181
6	191.4	19.1	32195	22	AAL04180
7	190.8	19.1	7677	23	ABV23205
8	190.8	19.1	7677	23	ABV29042

C	9	190.4	19.0	3074	21	AAF76994	Human CRFX ORF2549
C	10	189.8	19.0	10380	18	AA767164	Human alpha-N-acet
C	11	189.8	18.9	17792	22	AAS32727	Human genomic DNA
C	12	188.8	18.9	17792	22	AAS35099	Human cardiovascular
C	13	188.6	18.9	19616	22	ABA16094	Human nervous syst
C	14	188.6	18.8	19616	22	ABA16856	Human nervous syst
C	15	188.6	18.8	19616	22	AAL36855	Human musculoskele
C	16	188.6	18.8	19616	25	ABX59843	cDNA encoding nove
C	17	187.9	18.8	50000	21	AAA96366	Poly-orphic repea
C	18	187.9	18.7	4768	21	AAAC69134	Human ABC1 gene ex
C	19	187.9	18.7	11754	22	AAS04033	Human ABC1 gene, p
C	20	187.9	18.7	11754	24	ABL58149	Human ABCA1 transp
C	21	187.9	18.7	52616	22	AAK70459	Human immune/haema
C	22	187.9	18.7	52616	22	AAK78930	Human immune/haema
C	23	187.6	18.7	97835	24	ABK84796	Human cDNA differe
C	24	187.6	18.7	183999	22	AAF92831	Human ABC1 genomic
C	25	187	18.7	3417	22	AAH18467	Human cDNA sequenc
C	26	186.6	18.6	812	21	AAAC74416	Human secreted pro
C	27	186.6	18.6	6616	22	AAK81014	Human immune/haema
C	28	186.6	18.6	9469	22	AAK79514	Human immune/haema
C	29	186.6	18.6	17335	23	ABK42393	Genomic sequence #
C	30	186.6	19.6	19882	23	ABK42394	Genomic sequence #
C	31	186.6	18.6	79528	24	AAL50614	Human cancer statu
C	32	186.6	18.6	116840	25	ABQ77401	Human VWF DNA, Ho
C	33	186.6	18.6	1691080	24	ABX08336	Human phosphodiast
C	34	186.4	18.6	2233	22	AAK77256	Human immune/haema
C	35	186.4	18.6	13467	22	AAL05944	Human reproductive
C	36	186.4	18.6	13467	22	AAS27670	DNA encoding novel
C	37	186.4	18.6	13467	22	AAS27670	DNA encoding novel
C	38	186.4	18.6	13467	22	AAK85539	Human immune/haema
C	39	186.4	18.6	13467	23	ABL98508	Human testicular a
C	40	186.2	18.6	14491	25	ABZ09914	Human 5' and/or re
C	41	185.8	18.6	32190	22	AAS36709	Human cardiovascular
C	42	185.6	19.5	13058	22	AAS29194	Genomic sequence #
C	43	185.6	19.5	13058	22	AAK79325	Human immune/haema
C	44	185.6	18.5	13058	24	ABS68334	Human DNA-binding
C	45	185.6	18.5	62804	24	AAD39317	Human calcium/calr

ALIGNMENTS

RESULT 1

AAF84800
ID AAF84800 standard; DNA; 110000 BP.

XX AAF84800;

XX 09-JUL-2001 (first entry)

XX Nucleotide sequence of the human SPG4 gene.

XX Human; SPG4 gene; spastin; PSP-AD; gene therapy;
XX autosomal dominant familial spastic paraplegia; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 9932..102009

FT /tag= a

FT /note= "contains introns"

FT exon 9932..10471

FT /tag= b

FT /number= "1"

FT /tag= c

FT /number= "2"

FT /tag= d

FT /number= "2"

FT /number= "35748"

FT /tag= e

FT /number= "2"

FT

```
FT 35749..35832
FT /*tag= f
FT /number= "3"
FT 35833..45022
FT /*tag= g
FT /number= "3"
FT 45023..45118
FT /*tag= h
FT /number= "4"
FT 45119..60863
FT /*tag= i
FT /number= "4"
FT 60864..61051
FT /*tag= j
FT /number= "5"
FT 61052..61927
FT /*tag= k
FT /number= "5"
FT 61928..62061
FT /*tag= l
FT /number= "6"
FT 62062..62344
FT /*tag= m
FT /number= "6"
FT 62345..62438
FT /*tag= n
FT /number= "7"
FT 62439..73173
FT /*tag= o
FT /number= "7"
FT 73174..73248
FT /*tag= p
FT /number= "8"
FT 73249..74633
FT /*tag= q
FT /number= "8"
FT 74634..74705
FT /*tag= r
FT /number= "9"
FT 74706..82788
FT /*tag= s
FT /number= "9"
FT 82789..82864
FT /*tag= t
FT /number= "10"
FT 82865..83102
FT /*tag= u
FT /number= "10"
FT 83103..83194
FT /*tag= v
FT /number= "11"
FT 83195..83334
FT /*tag= w
FT /number= "11"
FT 83335..83414
FT /*tag= x
FT /number= "12"
FT 83415..88129
FT /*tag= y
FT /number= "12"
FT 88130..88172
FT /*tag= z
FT /number= "13"
FT 88173..89561
FT /*tag= aa
FT /number= "13"
FT 89562..89641
FT /*tag= ab
FT /number= "14"
FT 89642..91162
FT /*tag= ac
FT /number= "14"
FT 91163..91233
FT

FT 35749..35832
FT /*tag= ad
FT /number= "15"
FT 91234..93443
FT /*tag= ae
FT /number= "15"
FT 93444..93484
FT /*tag= af
FT /number= "16"
FT 93485..100599
FT /*tag= ag
FT /number= "16"
FT 100600..102009
FT /*tag= ah
FT /number= "17"
FT

XX PR2798138-A1.
XX
XX 09-MAR-2001.
XX
XX 03-SEP-1999; 99FR-0011097.
XX
XX 03-SEP-1999; 99FR-0011097.
XX
XX (CNRS : CNRS CENT NAT RECH SCI.
XX
XX Weissenbach J, Hazan J;
XX
XX WPI; 2001-283966/30.
XX
XX New human nucleic acid from the SPG4 gene, useful e.g. for diagnosis of
XX autosomal dominant familial spastic paraplegia and in drug screening .
XX
XX Claim 2; Page 45-106; 145pp; French.
XX
XX The present sequence represents a human SPG4 gene. The SPG4 gene encodes
XX a spastin polypeptide. Mutations in the SPG4 gene are responsible for
XX autosomal dominant familial spastic paraplegia. SPG4 polynucleotides,
XX and their fragments, are used to screen DNA banks for sequences that
XX encode spastin (particularly sequences in other mammals, specifically
XX mice); to identify SPG4 mutations, or other genetic anomalies,
XX particularly for diagnosis of autosomal dominant familial spastic
XX paraplegia (PSF-AD); to identify promoters and other regulatory elements
XX of the SPG4 gene; for detection and amplification; for recombinant
XX production of spastin; and for diagnostic genotyping of PSF-AD.
XX
XX Sequence 110000 BP; 30622 A; 21640 C; 22817 G; 34921 T; 3 other;
XX

Query Match 100.0%; Score 1001; DB 22; Length 110000;
Best Local Similarity 100.0%; Pred. No. 2e-171;
Matches 1001; Conservative C; Mismatches 3; Indels 0; Gaps 0;

QY 1 AATATAAAATGCATACCTTAGGCTGGGCAAGTGGCTACGCTGTGTAATCCGACACTT 60
DB |||||
DB 100000 AATATAAAATGCATACCTTAGGCTGGGCAAGTGGCTACGCTGTGTAATCCGACACTT 100059
QY 61 TGGGAGGCCAAGGTGGGGGATCACCTGAGGTTAGGAGTTAGGAGTTCAGGACCAACCTGGCCATCA 120
DB |||||
DB 100060 TGGGAGGCCAAGGTGGGGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCA 100119
QY 121 TGGCGAAACCCCTGTCTCTACTAAATAACAAAAATTAGTGGGCATGGAGGCATGTGCCT 180
DB |||||
DB 100120 TGGCGAAACCCCTGTCTCTACTAAATAACAAAAATTAGTGGGCATGGAGGCATGTGCCT 100179
QY 181 GTAATCCCAGGTACTCGGGAGGCTGAGACAGAGAAGTAATGCTTGAACCTTGGAGGGGGAGG 240
DB |||||
DB 100180 GTAATCCCAGGTACTCGGGAGGCTGAGACAGAGAAGTAATGCTTGAACCTTGGAGGGGGAGG 100239
QY 241 TTGCATATCTGAGTGGTGAATTTGTGATCTTTTTCCTCTCTCTCTCTCTCTCTCTCTCTCT 300
DB |||||
DB 100240 TTGCATATCTGAGTGGTGAATTTGTGATCTTTTTCCTCTCTCTCTCTCTCTCTCTCTCT 100299
QY 301 TTCTCTATAAATGATTGTGTTTGTGTTTATATATGGAAAAAATATTATGCTTTCAAATGTTA 360
DB |||||
```

Db 100100 TTTCTATAAATGATTGTCTTTTCTTTTATATATTTSSAAATAATATATGCTTTTCAAATGTTA 100159
QY 361 ATACCCATGAACACTAANCACAAAGTAAATATAATATATAGTATAGCATTTATTATAGGTTTCT 420
Db 100160 ATACCCATGAACACTAANCACAAAGTAAATATAATATATAGTATAGCATTTATTATAGGTTTCT 100419
QY 421 TGTGTACGAGATCAACATAGAAATATATTTTAAATGGCTGACATATTTTCTAAGATATAC 480
Db 100420 TGTGTACGAGATCAACATAGAAATATATTTTAAATGGCTGACATATTTTCTAAGATATAC 100479
QY 481 ATACACGTATATTTTATATACATTTAGANACAGACGATCATTTTAAATCCATCAT 540
Db 100482 ATACACGTATATTTTATATACATTTAGANACAGACGATCATTTTAAATCCATCAT 100539
QY 541 TCGTTAACACCATATACCTGTGTGATCTTGTATGTGATGGCTTTTAAATATCTAG 600
Db 100540 TCGTTAACACCATATACCTGTGTGATCTTGTATGTGATGGCTTTTAAATATCTAG 100599
QY 601 ATGAGAAATATTCGATATATCTGACITCACTGAATCTTGGAAATAAATAAGCAGCGTC 660
Db 100600 ATGAGAAATATTCGATATATCTGACITCACTGAATCTTGGAAATAAATAAGCAGCGTC 100659
QY 661 AGCCCTCAAACTTTAGAGCGGTACATACGTTGGAAACAGGACITTTGGAGATACCACTGT 720
Db 100662 AGCCCTCAAACTTTAGAGCGGTACATACGTTGGAAACAGGACITTTGGAGATACCACTGT 100719
QY 721 TAAGGAATATACCTTTGFAAACCTTGCAAGAACATTTTACTTTAAAGAGGAAACACAGATCT 780
Db 100720 TAAGGAATATACCTTTGFAAACCTTGCAAGAACATTTTACTTTAAAGAGGAAACACAGATCT 100779
QY 781 TCAATGAACGTATCGGTACAGAAACAGCGCTTAAGTTTACAGGACTTTTACAGTCTTAC 840
Db 100780 TCAATGAACGTATCGGTACAGAAACAGCGCTTAAGTTTACAGGACTTTTACAGTCTTAC 100839
QY 841 ATATTTGTCACCAAACTTTGAAGATGAACCCAGAAACAGACTTTAAACAAATATACAAATG 900
Db 100840 ATATTTGTCACCAAACTTTGAAGATGAACCCAGAAACAGACTTTAAACAAATATACAAATG 100899
QY 901 CAAATGTATATTTTGTGTGTTTAAAGCCTTGCTTGAATGGTCAAGTTATCCCAATGGAC 960
Db 100900 CAAATGTATATTTTGTGTGTTTAAAGCCTTGCTTGAATGGTCAAGTTATCCCAATGGAC 100959
QY 961 ACTAAGTTAGAGCACACAAACCAAACTTCTGCTGCTCTTT 1001
Db 100960 ACTAAGTTAGAGCACACAAACCAAACTTCTGCTGCTCTTT 101000

RESULT 2
AAAF84801
ID AAF84801 standard; cDNA; 3263 BP.
XX
AC AAF84801;
XX
D7 09-JUL-2001 (first entry)
XX
DE Nucleotide sequence of the human SPG4 polypeptide.
XX
KW Human: SPG4 gene; spastin; PSP-AD; gene therapy;
XX autosomal dominant familial spastic paraplegia; ss.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 126..1976
FT /*tag= a
FT /product= "spastin"
XX
PN FR2798:38-A1.
XX
PD 09-MAR-2001.
XX
PF 03-SEP-1999; 99FR-0011097.

FR 03-SEP-1999; 99FR-0011097.
XX
PA (CHPS) CNRS CENT NAT RECH SCI.
XX
PI Weissenhach J, Hazan J;
XX
DR WP1: 2001-283966/30.
DR P-PSDB; AAB68137.
XX
PT New human nucleic acid from the SPG4 gene, useful e.g. for diagnosis of
PT autosomal dominant familial spastic paraplegia and in drug screening -
XX
PS Claim 2: Page 106-108; 145pp, French.
XX
CC The present sequence encodes a human spastin polypeptide. Spastin is
CC encoded by the SPG4 gene. Mutations in the SPG4 gene are responsible for
CC autosomal dominant familial spastic paraplegia. SPG4 polynucleotides,
CC and their fragments, are used to screen DNA banks for sequences that
CC encode spastin (particularly sequences in other mammals, specifically
CC mice); to identify SPG4 mutations, or other genetic anomalies,
CC particularly for diagnosis of autosomal dominant familial spastic
CC paraplegia (PSP-AD); to identify promoters and other regulatory elements
CC of the SPG4 gene; for detection and amplification; for recombinant
CC production of spastin; and for diagnostic genotyping of PSP-AD.
SQ Sequence 3263 BP; 956 A; 664 C; 703 G; 940 T; 0 other;

Query Match 40.3%; Score 403; DB 22; Length 3263;
Best Local Similarity 100.0%; Pred. No. 8e-64;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 599 AGATGAGAAATATTCGATATATCTGACTTCACTGATCTCTTGAATAAATAAAGCGGCG 658
DB 1852 AGATGAGAAATATTCGATATATCTGACTTCACTGATCTCTTGAATAAATAAAGCGGCG 1911
QY 659 TCGCCCTCAAACTTTAGAGCGGTACATACGTTGGAAACAGGACTTTGGAGATACCACTG 718
DB 1912 TCGCCCTCAAACTTTAGAGCGGTACATACGTTGGAAACAGGACTTTGGAGATACCACTG 1971
QY 719 TTTAAGGAATATACCTTTGTAACCTTGCAGAACATTTTACTTTAAAGAGGAAACACAGAT 778
DB 1872 TTTAAGGAATATACCTTTGTAACCTTGCAGAACATTTTACTTTAAAGAGGAAACACAGAT 2031
QY 779 CTTGATGAACCTCATCGGTACAGAAACAGGCTTAAGTTTACAGGACTTTTACAGTCTT 838
DB 2032 CTTCAATGAACGTCACTGGGTACAGAAACAGGCTTAAGTTTACAGGACTTTTACAGTCTT 2091
QY 839 ACATATTTGTGACCAAACTTTGAAGATGAACCCAGAAACAGACTTTAAACAAATATACAA 898
DB 2092 ACATATTTGTGACCAAACTTTGAAGATGAACCCAGAAACAGACTTTAAACAAATATACAA 2151
QY 899 TCGAAATGTAAATTTTGTGTGTTTAAAGCCTTGCTTGAATGGTCAAGTTATCCCAATGG 959
DB 2152 TCGAAATGTAAATTTTGTGTGTTTAAAGCCTTGCTTGAATGGTCAAGTTATCCCAATGG 2211
QY 959 ACATTAAGTTAGAGCACACAAACCAAACTTCTGCTGCTCTTT 1001
DB 2212 ACATTAAGTTAGAGCACACAAACCAAACTTCTGCTGCTCTTT 2254

RESULT 3
ABQ54772
ID ABQ54772 standard; cDNA; 4152 BP.
XX
AC ABQ54772;
XX
D7 22-AUG-2002 (first entry)
XX
DE Human ovarian antigen HNTTB23 cDNA, SEQ ID NO:652.
XX
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;

KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
KW inflammatory condition; immune disorder; blood disorder;
KW cardiovascular disorder; respiratory disorder; neurological disorder;
KW gastrointestinal disorder; urinary system disorder; drug screening;
KW gene therapy; chromosome mapping; forensic analysis;
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW anti-inflammatory; gynaecological; reproductive; gene; ss.
XX
OS Homo sapiens.
XX WC200200677-A1.
XX
PD 03-JAN-2002.
XX
PF 07-JUN-2001; 2001WC-US18569.
XX
PR 07-JUN-2000; 2000US-209467P.
XX
PA (HUNA-) HUMAN GENOME SCI INC.
XX
FI Birse CE, Rosen CA;
XX WPI; 2002-147878/19.
DR P-PSDE; ABP41695.
XX
PT isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT useful in the prevention, treatment and diagnosis of cancer (e.g.
PT ovarian cancer), immune disorders, cardiovascular disorders and
PT neurological diseases -
XX
PS Claim 1; SEQ ID NO 652; 29222BP; English.
XX
CC The invention relates to 2175 novel human ovarian antigens (ABP41654-
CC ABP43228) and to cDNAs encoding them (ABQ5431-ABQ56305), and also
CC encompasses polypeptides 90% identical and polynucleotides 95% identical
CC to the sequences of the invention. The invention additionally relates to
CC recombinant vectors and host cells comprising human ovarian antigen
CC polynucleotides, antibodies against human ovarian antigens, and the use
CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
CC treating, prognosing or preventing various ovary and/or breast-related
CC disorders. Such conditions include ovarian cancer and breast cancer, and
CC metastatic tumours of ovarian or breast origin, reproductive system
CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine
CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
CC vaginitis), immune disorders (e.g., congenital and acquired
CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
CC respiratory disorders, neurological disorders, gastrointestinal disorders
CC and urinary system disorders. Ovarian antigen polypeptides and
CC polynucleotides may also be used in screening for compounds which
CC modulate ovarian antigen expression or activity. The polynucleotides may
CC further be used for gene therapy, chromosome mapping, in the
CC identification of individuals and in forensic analysis, and the
CC polypeptides may be used as food additives or to prepare antibodies
CC useful in disease diagnosis, drug targeting and phenotyping. The present
CC sequence represents cDNA encoding a human ovarian antigen of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 4152 BP; 1331 A; 663 C; 777 G; 1376 T; 5 other;

Query Match 40.3%; Score 403; DB 24; Length 4152;
Best Local Similarity 100.0%; Pred. No. 8,1e-64;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 599 AGATGAGAAATATTGATTATCTGACTTCACCTGACCTGAGTAAATTAACCGCAGG 659
D5 853 AGATGAGAAATATTGATTATCTGACTTCACCTGACCTGAGTAAATTAACCGCAGG 912

QY 659 TCAGCCCTCAAACTTTAGAGGCTACATACGTTGGAAACAAGGACTTTGGAGATACCACTG 718
D5 913 TCAGCCCTCAAACTTTAGAGGCTACATACGTTGGAAACAAGGACTTTGGAGATACCACTG 972
QY 719 TTTAAGGAAATACCTTTGTAACCTGCAGAACATTTTACTTAAAGAGGAAACACAAAGAT 778
D5 973 TTTAAGGAAATACCTTTGTAACCTGCAGAACATTTTACTTAAAGAGGAAACACAAAGAT 1032
QY 779 CTTCAATGAACGTCATCGGCTACAGAAACAGCTTAAGTTTACAGGACTTTTAGAGGCTT 838
D5 1033 CTTCAATGAACGTCATCGGCTACAGAAACAGCTTAAGTTTACAGGACTTTTAGAGGCTT 1092
QY 839 ACATAATTTGTCACCAAACTTGAAGATGAACAGAAACAGACTTAAACAAATATACAA 898
D5 1093 ACATAATTTGTCACCAAACTTGAAGATGAACAGAAACAGACTTAAACAAATATACAA 1152
QY 899 TGCATATGTAATTTTGTGTTAAGGCTTGCCTTGATGGTCACAGTTATCCCAATGG 958
D5 1153 TGCATATGTAATTTTGTGTTAAGGCTTGCCTTGATGGTCACAGTTATCCCAATGG 1212
QY 959 ACATAATGTTAGGACACAAACAACTGATTTGCTGCTTCCTT 1001
D5 1213 ACATAATGTTAGGACACAAACAACTGATTTGCTGCTTCCTT 1255

RESULT 4

AAS17746

ID AAS17746 standard; DNA; 3143 BP.

XX AAS17746;

XX AAS17746;

DT 26-FEB-2002 (first entry)

XX Human genomic clone for P2Y-like G protein-coupled receptor.

XX Human; ds; P2Y-like G protein-coupled receptor; GPCR; COPD;

XX chronic obstructive pulmonary disease; nervous system disease;

XX Parkinson's disease; multiple sclerosis; dementia; stroke;

XX Alzheimer's disease; benign prostatic hyperplasia; urinary incontinence;

XX bacterial infection; fungal infection; protozoan infection;

XX viral infection; pain; cancer; anorexia; bulimia; asthma;

XX acute heart failure; hypertension; hypertension; osteoporosis; diabetes;

XX angina pectoris; myocardial infarction; ulcer; inflammation; allergy;

XX psychotic disorder; neurological disorder; anxiety; schizophrenia;

XX manic depression; delirium; severe mental retardation; dyskinesia.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 520..2498

FT /*tag= a

FT /product= "P2Y-like GPCR"

FT

XX WC200185754-A2.

XX 15-NOV-2001.

XX 09-MAY-2001; 2001WO-EP05244.

XX 11-MAY-2000; 2000US-203582P.

XX 21-FEB-2001; 2001US-269857P.

XX (FARB) BAYER AG.

XX Ramakrishnan S;

XX WPI; 2002-075242/10.

XX P-PSDE; AAS11251.

XX New polynucleotides for producing P2Y-like G protein-coupled receptors

XX (GPCR) that are used for screening inhibitors or regulators of human

XX P2Y-like GPCR, especially useful for treating pain, cancer or

XX neurological disorders -

PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237043.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0243960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246539.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246612.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251010.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465570/50.
XX
PT isolated nucleic acid molecule encoding a reproductive system antigen
PT is used in preventing, treating or ameliorating a medical condition
XX
PS Disclosure; SEQ ID NO 6869; 1297pp + Sequence listing; English.
XX

CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a genomic sequence encoding a
CC protein of the invention.
XX
SQ Sequence 30573 BP; 7764 A; 7219 C; 7253 G; 8337 T; 0 other;
Query Match 19.1%; Score 191.4; DS 22; Length 30573;
Best Local Similarity 90.7%; Pred. No. 11e-25;
Matches 204; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
QY 21 AGGCTGGGCAAGTGGTCAAGTCTGTAAATCCAGACCTTTGGGAGGCCAAAGGTGGCGG 80
DB 6984 AGGCTGGGCAAGTGGTCAAGTCTGTAAATCCAGACCTTTGGGAGGCCAAAGGTGGCGG 7043
QY 51 ATCACCTGAGTTAGGAGTTCAGGACCAACCTGGCCATCATATGCCAGCTACTCTCTAC 140
DB 7044 ATCATCTGAGGTCAGGAGTTCAGGACCAACCTGGCCATCATATGCCAGCTACTCTCTAC 7103
QY 141 TAAATAATACAAAATTAGCTGGGATGGAGGCTGTGCTGTAAATGCCAGCTACTCTCTAC 200
DB 7104 TAAATAATACAAAATTAGCTGGGATGGAGGCTGTGCTGTAAATGCCAGCTACTCTCTAC 7163
QY 201 GGCTGAGACAGAGAAATTGCTTGAACCTTGGAGGGGAGGAGGTTGCA 245
DB 7164 GGCTGAGGAGGAGAAATTGCTTGAACCTTGGAGGGGAGGAGGTTGCA 7208
RESULT 6
AAL04180
ID AAL04180 standard; DNA; 32195 BP.
XX
AC AAL04180;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 6868.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy; ds.
XX
CS Homo sapiens.
XX
PN W0200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2000; 2001WO-US01339.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0196123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.

PR	14-AUG-2000;	2000JUS-0225567.
PR	14-AUG-2000;	2000JUS-0225568.
PR	14-AUG-2000;	2000JUS-0225570.
PR	14-AUG-2000;	2000JUS-0225571.
PR	14-AUG-2000;	2000JUS-0225577.
PR	14-AUG-2000;	2000JUS-0225581.
PR	14-AUG-2000;	2000JUS-0225588.
PR	14-AUG-2000;	2000JUS-0225599.
PR	18-AUG-2000;	2000JUS-0225629.
PR	22-AUG-2000;	2000JUS-0226681.
PR	22-AUG-2000;	2000JUS-0226686.
PR	22-AUG-2000;	2000JUS-0227182.
PR	23-AUG-2000;	2000JUS-0227539.
PR	30-AUG-2000;	2000JUS-0228924.
PR	01-SEP-2000;	2000JUS-0229287.
PR	01-SEP-2000;	2000JUS-0229343.
PR	01-SEP-2000;	2000JUS-0229344.
PR	01-SEP-2000;	2000JUS-0229345.
PR	05-SEP-2000;	2000JUS-0229539.
PR	05-SEP-2000;	2000JUS-0229533.
PR	06-SEP-2000;	2000JUS-0230437.
PR	06-SEP-2000;	2000JUS-0230438.
PR	08-SEP-2000;	2000JUS-0231242.
PR	08-SEP-2000;	2000JUS-0231243.
PR	08-SEP-2000;	2000JUS-0231244.
PR	08-SEP-2000;	2000JUS-0231413.
PR	08-SEP-2000;	2000JUS-0231414.
PR	08-SEP-2000;	2000JUS-0232580.
PR	08-SEP-2000;	2000JUS-0232581.
PR	12-SEP-2000;	2000JUS-0231968.
PR	14-SEP-2000;	2000JUS-0232397.
PR	14-SEP-2000;	2000JUS-0232398.
PR	14-SEP-2000;	2000JUS-0232399.
PR	14-SEP-2000;	2000JUS-0232400.
PR	14-SEP-2000;	2000JUS-0232401.
PR	14-SEP-2000;	2000JUS-0233063.
PR	14-SEP-2000;	2000JUS-0233064.
PR	14-SEP-2000;	2000JUS-0233065.
PR	21-SEP-2000;	2000JUS-0234273.
PR	21-SEP-2000;	2000JUS-0234274.
PR	25-SEP-2000;	2000JUS-0234397.
PR	25-SEP-2000;	2000JUS-0234398.
PR	26-SEP-2000;	2000JUS-0235484.
PR	27-SEP-2000;	2000JUS-0235834.
PR	27-SEP-2000;	2000JUS-0235836.
PR	28-SEP-2000;	2000JUS-0236327.
PR	29-SEP-2000;	2000JUS-0236367.
PR	29-SEP-2000;	2000JUS-0236368.
PR	29-SEP-2000;	2000JUS-0236369.
PR	29-SEP-2000;	2000JUS-0236370.
PR	02-OCT-2000;	2000JUS-0236802.
PR	02-OCT-2000;	2000JUS-0237037.
PR	02-OCT-2000;	2000JUS-0237038.
PR	02-OCT-2000;	2000JUS-0237539.
PR	02-OCT-2000;	2000JUS-0237540.
PR	13-OCT-2000;	2000JUS-0239935.
PR	13-OCT-2000;	2000JUS-0239937.
PR	20-OCT-2000;	2000JUS-0240900.
PR	20-OCT-2000;	2000JUS-0241221.
PR	20-OCT-2000;	2000JUS-0241795.
PR	20-OCT-2000;	2000JUS-0241786.
PR	20-OCT-2000;	2000JUS-0241787.
PR	20-OCT-2000;	2000JUS-0241808.
PR	20-OCT-2000;	2000JUS-0241809.
PR	20-OCT-2000;	2000JUS-0241826.
PR	01-NOV-2000;	2000JUS-0244617.
PR	08-NOV-2000;	2000JUS-0245474.
PR	08-NOV-2000;	2000JUS-0246475.
PR	08-NOV-2000;	2000JUS-0246476.
PR	08-NOV-2000;	2000JUS-0246477.
PR	08-NOV-2000;	2000JUS-0246478.
PR	08-NOV-2000;	2000JUS-0246523.
PR	08-NOV-2000;	2000JUS-0246524.
PR	08-NOV-2000;	2000JUS-0246525.

PR	08-NOV-2000	2000US-0246526
PR	08-NOV-2000	2000US-0246527
PR	08-NOV-2000	2000US-0246528
PR	08-NOV-2000	2000US-0246532
PR	08-NOV-2000	2000US-0246609
PR	08-NOV-2000	2000US-0246610
PR	08-NOV-2000	2000US-0246611
PR	08-NOV-2000	2000US-0246613
PR	17-NOV-2000	2000US-0249217
PR	17-NOV-2000	2000US-0249218
PR	17-NOV-2000	2000US-0249219
PR	17-NOV-2000	2000US-0249220
PR	17-NOV-2000	2000US-0249221
PR	17-NOV-2000	2000US-0249222
PR	17-NOV-2000	2000US-0249223
PR	17-NOV-2000	2000US-0249214
PR	17-NOV-2000	2000US-0249215
PR	17-NOV-2000	2000US-0249216
PR	17-NOV-2000	2000US-0249217
PR	17-NOV-2000	2000US-0249218
PR	17-NOV-2000	2000US-0249244
PR	17-NOV-2000	2000US-0249245
PR	17-NOV-2000	2000US-0249264
PR	17-NOV-2000	2000US-0249265
PR	17-NOV-2000	2000US-0249297
PR	17-NOV-2000	2000US-0249299
PR	17-NOV-2000	2000US-0249300
PR	01-DEC-2000	2000US-0250160
PR	01-DEC-2000	2000US-0250191
PR	05-DEC-2000	2000US-0251030
PR	05-DEC-2000	2000US-0251988
PR	05-DEC-2000	2000US-0256719
PR	06-DEC-2000	2000US-0251479
PR	08-DEC-2000	2000US-0251856
PR	08-DEC-2000	2000US-0251868
PR	08-DEC-2000	2000US-0251869
PR	08-DEC-2000	2000US-0251969
PR	11-DEC-2000	2000US-0251990
PR	05-JAN-2001	2000US-0254097
PR	05-JAN-2001	2000US-0253678

Query Match 19.1%; Score 191.4; DB 22; Length 32195;
Best Local Similarity 90.7%; Pred. No. Size=25;

QY	21	AGGCTGGGCAAGTGCTCACGCTGTGTAATCCAGCACCTTGGAGGGCCAAAGGTGGGGCGG	50
DB	12766	AGGCTGGGCAAGTGCTCACACCTGTAATCCAGCACCTTGGAGGTCGAGGTAGGTGG	12825
QY	91	ATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCAAAACCCCTGCTCTAC	140
DB	12826	ATCATCTGAGGTCAGGAGTTCAGACCAAGCCTGGCCAAACATGCTGAAACCCCTGCTCTAC	12885
QY	141	TAAATAACAATAATTAGCTGGGCATGGAGGCATGTGCCCTGTAAATCCAGTACTCGGGA	200

```

Db      12886 TAAAAATACAAATAGCTGGGATGGTGGACCTGCTGTAAATCCAGCTACTGGGA 12945
QY      201 GGCTGAGACAGAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
Db      12946 GGCTGAGGACAGAGAATTGCTTGAACCTGGGAGGTTGGAGTTGCA 12990

RESULT 7
ABV23205
ID      ABV23205 standard; cDNA: 7677 BP.
XX      AC      ABV23205;
XX      DT      16-SEP-2002 (first entry)
XX      DE      Human prostate expression marker cDNA 23196.
XX      KW      Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX      KW      pharmacogenomic marker; gene; ss.
XX      OS      Homo sapiens.
XX      PN      WC20016C860-A2.
XX      PD      23-AUG-2001.
XX      PF      20-FEB-2001; 2001WC-US05171.
XX      PR      17-FEB-2000; 2000US-183319P.
XX      PR      16-MAR-2000; 2000US-189862P.
XX      PR      25-MAY-2000; 2000US-207454P.
XX      PR      09-JUN-2000; 2000US-211314P.
XX      PR      18-JUL-2000; 2000US-219007P.
XX      PR      13-DEC-2000; 2000US-255281P.
XX      (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX      PI      Schlegel R, Endege WO, Morahan JE;
XX      DR      WPI; 2001-662795/76.
XX      PT      Novel isolated nucleic acid molecule associated with cancerous state of
XX      PT      prostate cells and correlating with presence of prostate cancer, useful
XX      PT      for detecting presence of prostate cancer, stage of prostate cancer.
XX      PS      Claim 1; Page 4173-4175; 11750pp; English.
XX      CC      The invention relates to an isolated nucleic acid molecule (I) comprising
XX      CC      a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX      CC      specification or its complement. (II) is useful for:
XX      CC      (a) assessing whether a patient is afflicted with prostate cancer;
XX      CC      (b) monitoring the progression of prostate cancer in a patient;
XX      CC      (c) assessing the efficacy of a test compound to inhibit prostate
XX      CC      cancer in a patient;
XX      CC      (d) assessing the efficacy of a therapy for inhibiting prostate cancer
XX      CC      in a patient;
XX      CC      (e) selecting a composition for inhibiting prostate cancer in a patient;
XX      CC      (f) assessing the prostate cell carcinogenic potential of a compound;
XX      CC      (g) determining whether prostate cancer has metastasized in a patient;
XX      CC      (h) assessing the aggressiveness or incidence of prostate cancer in a
XX      CC      patient;
XX      CC      (i) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX      SQ      Sequence 7677 BP; 2389 A; 1642 C; 1794 G; 1846 T; 6 other;

Query Match      19.1%; Score 190.8; DB 23; Length 7677;
Best Local Similarity 60.7%; Pred. No. 1.3e-25;
Matches 368; Conservative 0; Mismatches 227; Indels 11; Gaps 3;

QY      1 AATATAAAAATGCTACTTTAGCTGGGCAAGTGGCTCAGCTCTGTAAATCCAGCACTT 60
Db      6836 AATTATTAAAGATTAAATCTCGGCTGGGGCGGTGGCTCAGCTCTGTAAATCCAGCACTT 6895

```

```

QY      61 TGGAGGCCCAAGGTGGGCGGATCAGCTGAGGTAGGAGTTTACAGCAACCACTGGCCATCA 120
Db      6896 TGGGAAGCCCAAGGTGGGCGGATCAGCTTCCAGGTAGGAGTTTGGAGACCAGCTTGGCCAACA 6955
QY      121 TGGGAAACCCCTGTCTCTACTATAAAATACAAAAATAGCTGGGCATGGAGGCATGTGCT 180
Db      6956 TGGGAAACCCCTGTCTCTACTATAAAATACAAAAATAGCCGGCGGTGTGGCCTTGTGCT 7015
QY      181 GTAATCCAGCTACTCGGGAGGCTGAGACAGAGAATTCCTTGAACCTTGGAGGGGGAGG 240
Db      7016 GTAATCCAGCTACTCGGGAGGCTGAGGCCAGCAAGATCGATTAAACCTGGAGGGCGGAGA 7075
QY      241 TTGC----ATATCTCAGTGGTGAATTTGTGATTTCTTTTCTCTCTTTTCTCTCTTTTCT 296
Db      7076 TTGCATGAGATCGAACCACTGCACTCCAGCTCCAGCTGGGTGACAGAGAGAGACACTGCCCTTG 7135
QY      297 AACTTTTCTATAAATGATTGTTGTTTGTATATTTTATATTTGGAAAAATATATGCTTTTCAAA 356
Db      7136 AAAAAAAGAAATCTCACTCACTATCTAGAGAGGATTGTCAGAAATATTCAGCATTCAGGT 7195
QY      357 GTTAATACCTATGAACCTAAACACAGAAATATAAATATATTAGTATAGCATTTATTAAAGT 416
Db      7196 CTTGAACCTTTGATTATGCAAAAGAGGTATATAAATAATATTTCATTATGATTCAAGTTT 7255
QY      417 TTCTTGTGTAGCAGATCAACATAGAAAAATATATTAAATGGCTGACATTAATTTCTAAGA 476
Db      7256 TTAAGGCTTTGCAGCT---TCTATAAGTGTCTCAGATGCCACTAGATAATTTTAAAGC 7312
QY      477 ATACATACAGGTATATTTTATAACATTA---AGAACAGAGCAGCATCATTTTAAAT 532
Db      7313 ATCATATTAGAAATACCTTTAAGAGACCTTATATAAGAAATAGAAGATTGTTCAATTTTAC 7372
QY      533 CCATCATTTGTTTAAACCACCATATACCTGTGTGATCATTTGTATTGTGATGCTTTTAA 592
Db      7373 AGAGGATTGGTTTCATTAAGACCCAGATTCTGTAAATTTTCACTTGTGAATTTAGTTAA 7432
QY      593 AAATCT 598
Db      7433 ACATA 7438

```

```

RESULT 8
ID      ABV29042
XX      AC      ABV29042;
XX      DT      16-SEP-2002 (first entry)
XX      DE      Human prostate expression marker cDNA 29033.
XX      KW      Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX      KW      pharmacogenomic marker; gene; ss.
XX      OS      Homo sapiens.
XX      PN      WC20016C860-A2.
XX      PD      23-AUG-2001.
XX      PF      20-FEB-2001; 2001WC-US05171.
XX      PR      17-FEB-2000; 2000US-183319P.
XX      PR      16-MAR-2000; 2000US-189862P.
XX      PR      25-MAY-2000; 2000US-207454P.
XX      PR      09-JUN-2000; 2000US-211314P.
XX      PR      18-JUL-2000; 2000US-219007P.
XX      PR      13-DEC-2000; 2000US-255281P.
XX      (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX      PI      Schlegel R, Endege WO, Morahan JE;

```



```
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 3074 BP; 654 A; 851 C; 847 G; 720 T; 2 other;

Query Match          19.0%; Score 190.4; DB 21; Length 3074;
Best Local Similarity 87.1%; Pred. No. 1.5e-25;
Matches 209; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 6 AAAAAATGCATCTTTAGGCTGGGCAAGTGGCTACGCTCTGTATATCCAGCACTTTGGGA 65
DB 1593 AAAAGTGCCTTTAGAGGCTGGGCATGGTGGCTCACGCTTGTATCCAGCACTTTGGGA 1534

QY 66 GGCCAAAGTGGGCGGATCACCTGAGGTAGGATTGAGGACCAACCTGGCCATCATGGCG 125
DB 1533 GGCCAAAGTGGTGGATCACCTGAGGTCAGGAGTTCAAGACCAGGCTGGCCAAACATGGTG 1474

QY 126 AAACCCGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCTGTAA 185
DB 1473 AAACCCGTCTCTACTAAAAATACAAAAATTAGCCAGGCATGGTACAGGCTGCTGTAA 1414

QY 186 CCCAGCTACTCGGGAGGCTGAGACAGAGAAATTCCTGAACCTTGGAGGGGAGGTTGCA 245
DB 1413 CCCAGCTACTCGGGAGGCTGAGGCGAGGAGAAATTCCTGAATCTGGAAGCGAGGTTGCA 1354

RESULT 10
AAT67164
ID AAT67164 standard; cDNA; 10380 BP.
XX
AC AAT67164;
XX
CT 20-AUG-1997 (first entry)
XX
DE Human alpha-N-acetylglucosaminidase gene.
XX
KW Alpha-N-acetylglucosaminidase; mucopolysaccharidosis type IIIB;
KW gene therapy; enzyme replacement therapy; diagnosis: ss.
XX
OS Homo sapiens.
XX
FH Location/Qualifiers
FT 5'UTR 1..989
FT exon 990..1372
FT intron 1373..2114
FT exon 2115..2262
FT intron 2263..3055
FT intron 3056..3202
FT intron 3203..3386
FT exon 3387..3472
FT intron 3473..5666
FT exon 5667..5923
FT intron 5924..7744
FT exon 7745..8955
FT intron 8956..10380
FT 3'UTR
```

```
FT /*tag= *
XX WC9719177-A1.
XX 29-MAY-1997.
XX 22-NOV-1996; 96WO-AU00747.
XX 23-NOV-1995; 95AU-C006748.
XX (WONE-) WOMEN'S & CHILDREN'S HOSPITAL.
XX Anson DS, Bianch L, Hopwood JJ, Scott H, Weber B;
XX MPI; 1997-298114/27.
XX P-PSDB; AAW18017.
XX
XX Nucleic acid encoding mammalian alpha-N-acetylglucosaminidase -
XX used for the diagnosis and treatment of mucopolysaccharidosis type
XX IIIB; also used in gene therapy
XX
XX Claim 8; Page 54-61; 79pp; English.
XX
XX A genomic DNA molecule (AAT67164) includes 6 exons that code for
XX human alpha-N-acetylglucosaminidase (AAW18017), an enzyme that can
XX hydrolyse the terminal alpha-N-acetylglucosamine residues at the
XX non-reducing terminus of fragments of heparan sulphate and heparin.
XX It was isolated by hybridisation of a human chromosome 17 library.
XX A cDNA clone (AAT67163) coding for the enzyme has also been isolated.
XX The isolated gene or cDNA, and primers/probes based on them or
XX their complementary strands, can be used to investigate, diagnose
XX and treat alpha-N-acetylglucosaminidase deficiency, for example in
XX patients suffering from mucopolysaccharidosis type IIIB.
XX Administration is by oral, i.v., i.p., enzyme replacement therapy,
XX gene therapy or other routes.
XX
SQ Sequence 10380 BP; 2210 A; 2953 C; 2851 G; 2366 T; 0 other;

Query Match          19.0%; Score 189.8; DB 15; Length 10380;
Best Local Similarity 86.7%; Pred. No. 2e-25;
Matches 209; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 5 TAAAAATGCATCTTTAGGCTGGGCAAGTGGCTACGCTCTGTATATCCAGCACTTTGGG 64
DB 3540 TAAAACTAAGCTCTGGGCGGGGGGAGTGGCTACGCTGTATATCCAGCACTTTGGG 3599

QY 65 AGGCCAAGGTGGGCGGATCACCTGAGGTAGGATTAGGACCAACCTGGCCATCATGGC 124
DB 3600 AGGCCGAGTTGGGCGGATCACCTGAGGTACAGGAGTTTCAGACCCAGCTGGCCACATGGT 3659

QY 125 GAAACCCGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCTGTAA 184
DB 3660 GAAACCCGTCTCTACTAAAAATATATAAAATTAGCTGGGCATGGTGGCAGCGGCTGTAA 3719

QY 185 TCCCAGCTACTCGGGAGGCTGAGACAGAGAAATTGCTTGAACCTTGGAGGGGAGGTTGC 244
DB 3720 TCCCATCTACTTAGGAGGCTGAGACAGAGAAATTGCTTAAACCTTGGAGGCAGACGTTGC 3779

QY 245 A 245
DB 3750 A 3780

RESULT 11
AAS32727
ID AAS32727 standard; DNA; 17792 BP.
XX
AC AAS32727;
XX
DT 17-DEC-2001 (first entry)
XX
DE Human genomic DNA for novel endocrine antigen, SEQ ID No 681.
XX
```

KW Human; endocrine antigen; ds; cytostatic; antifertility; antidiabetic;
KW thyroid-active; adrenal-active; androgenic; gastric gene therapy;
KW antisense-therapy; antibody; endocrine disorder; hormone imbalance;
KW reproductive disorder; endocrine cancer; pancreatic disorder;
KW diabetes mellitus; adrenal gland disorder; hirsutism; thyroid disorder;
KW hyperthyroidism; hypothalamic disorder; vanishing testes syndrome;
XX

CS Homo sapiens.

XX

PN WO200155319-A2.

XX

PD 02-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01335.

XX

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226581.

PR 22-AUG-2000; 2000US-0226588.

PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0228287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237038.

PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0238935.

PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.

PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.

PR 17-NOV-2000; 2000US-0249208.

PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.

PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.

PR 17-NOV-2000; 2000US-0249213.

PR 17-NOV-2000; 2000US-0249214.

PR 17-NOV-2000; 2000US-0249215.

PR 17-NOV-2000; 2000US-0249216.

PR 17-NOV-2000; 2000US-0249217.

PR 17-NOV-2000; 2000US-0249218.

PR 17-NOV-2000; 2000US-0249244.

PR 17-NOV-2000; 2000US-0249245.

PR 17-NOV-2000; 2000US-0249264.

PR 17-NOV-2000; 2000US-0249265.

PR 17-NOV-2000; 2000US-0249297.

PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.

PR 01-DEC-2000; 2000US-0250160.

PR 01-DEC-2000; 2000US-0250391.

PR 05-DEC-2000; 2000US-0251030.

PR 05-DEC-2000; 2000US-0251988.

PR 05-DEC-2000; 2000US-0256719.

PR 06-DEC-2000; 2000US-0251479.

PR 08-DEC-2000; 2000US-0251856.

PR 08-DEC-2000; 2000US-0251868.

PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-457726/49.
XX
PT Isolated polypeptide for treating, preventing and prognosing disorders
PT related to the endocrine system including endocrine disorders,
PT reproductive disorders, and gastrointestinal disorders and also for
PT testing and detection e.g. diagnosis -
XX
PS Disclosure; SEQ ID No 681; 55app; English.
XX
CC The invention relates to cDNAs encoding novel human endocrine
CC antigens or a fragment having biological activity, a domain, an epitope,
CC full length protein, variant, allelic variant or a species homologue of
CC the cDNA/antigen. The DNAs and polypeptides are useful for preventing,
CC treating or ameliorating a medical condition when administered
CC (e.g. by gene therapy or antisense-therapy). Identifying mutations in
CC the genes coding for the antigens is useful for diagnosing a pathological
CC condition or a susceptibility to a pathological condition. The DNAs,
CC antigens and antibodies raised against the antigens useful for treating,
CC preventing and/or prognosing disorders related to the endocrine system
CC or hormone imbalance or reproductive disorders, cancers of endocrine
CC tissues, disorders of the pancreas (e.g. diabetes mellitus), the adrenal
CC glands (e.g. hirsutism), ovaries, the thyroid (e.g. hyperthyroidism), the
CC hypothalamus and testes (e.g. vanishing testes syndrome), many examples
CC of diseases and disorders are given in the specification. The present
CC sequence is genomic DNA fragment form a gene encoding an endocrine
CC antigen of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pat_sequences.
XX
SQ Sequence 17792 BP; 4372 A; 4526 C; 4413 G; 4381 T; 0 other;

Query Match 18.9%; Score 186.8; DR 22; Length 17792;
Best Local Similarity 86.7%; Pred. No 3.1e-25;
Matches 208; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 6 AAAAAATGCTACTTTAGGCTGGGCAAGTGGCTCAGCTGTGTAAATCCAGCACCTTTGGGA 65
D5 13401 AAAAAATGCTACTTTAGGCTGGGCAAGTGGCTCAGCTGTGTAAATCCAGCACCTTTGGGA 13460

QY 66 GGCCCAAGGTGGGCGGATCAGCTGAGGTAGGATGAGGACCAACCTGGCCATCATGGCG 125
D5 13461 GGCCCAAGGTGGGCGGATCAGCTGAGGTAGGATGAGGACCAACCTGGCCATCATGGCG 13520

QY 126 AAACCCCTGTCTCTACTTAAATACAAATAGCTGGGATGGAGGCAATGTCCTGTAT 185
D5 13521 AAACCCCTGTCTCTACTTAAATACAAATAGCTGGGATGGAGGCAATGTCCTGTAT 13580

QY 186 CCCAGCTACTCGGAGGCTGACACAGAGAAATGCTTGAACCTTGGAGGGGAGGTGCA 245
D5 13581 CCCAGCTACTCGGAGGCTGACACAGAGAAATGCTTGAACCTTGGAGGGGAGGTGCA 13640

RESULT 12
AAS36099
ID AAS36099 standard; DNA; 17792 BP.
XX
AC AAS36099;
XX
DT 17-DEC-2001 (first entry)
XX
DE Human cardiovascular system antigen genomic DNA SEQ ID No 1539.

XX Cardiovascular system antigen; human; mouse; rabbit; goat; horse; cat;
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic; dog;
KW antirheumatic; antiproliferative; cytostatic; cardiac; neuroprotective;
KW cerebroprotective; necrotic; antibacterial; virucide; fungicide; cancer;
KW ophthalmological; vulnary; gene therapy; autoimmune disease; neoplasia;
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW cerebrovascular disorder; nervous system disorder; bacterial infection;
KW fungal infection; viral infection; ocular disorder; endocrine disorder;
KW gastrointestinal disorder; renal disorder; respiratory disorder;
KW wound healing; skin aging; organ transplantation; tissue regeneration;
KW anti-infertility.
XX Homo sapiens.
OS
XX WO200155321-A2.
PN
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2000; 200-WO-US01340.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 24-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235936.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.

PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259673.
XX
PA (HUMA:) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-451930/48.
XX
XX New cardiovascular system related polynucleotides and polypeptides,
XX useful for diagnosing, treating and/or preventing disorders of the
XX cardiovascular system -
XX
PS Claim 1; SEQ ID No 1599; 674pp; English.
XX
XX Sequences AAS35741-AAS36942 represent genomic DNA molecules, which encode
XX the cardiovascular system antigen polypeptides of the invention.
XX Cardiovascular system antigens and their associated polynucleotides are
XX useful in the diagnosis, treatment and prevention of various types of
XX disorders in e.g. humans, mice, rabbits, goats, horses, cats, dogs,
XX chickens or sheep. A pathological condition can be determined by
XX detecting the presence or absence of a mutation in a cardiovascular
XX system antigen polynucleotide. The treatable disorders include autoimmune
XX diseases such as rheumatoid arthritis, hyperproliferative disorders such
XX as neoplasms of the breast or liver, cardiovascular disorders such as
XX cardiac arrest, cerebrovascular disorders such as cerebral ischaemia,
XX nervous system disorders such as Alzheimer's disease, infections caused
XX by bacteria, viruses and fungi, ocular disorders such as corneal
XX infection, endocrine disorders such as premature labour and infertility,
XX gastrointestinal disorders such as Crohn's disease, renal disorders such
XX as glomerulonephritis and respiratory disorders such as asthma and
XX pleurisy. The polypeptides can also be used to aid wound healing, to
XX prevent skin aging due to sunburn, to maintain organs before
XX transplantation, to regenerate tissues and in chemotaxis.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.

Query Match 12.9%; Score 188.8; EB 22; Length 17792;
Best Local Similarity 86.7%; Pred. NO. 3.1e-25;
Matches 208; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 5 AAAATGCATACCTTAGGCTGGGCAAGTGGCTCAGCTCTGTATCCACGACCTTTGGGA 65
DB 13401 AAAAATCACTCTTTCGGCTGGGCGCAGTGGCTCAGCTCTGTATCCACGACCTTTGGGA 13460
QY 66 GGGCAAGGTGGGCGGATCACCTCAGGTTAGGATTCAGGACCACTGGGCAATGATGGG 125
DB 13461 GGGCAAGGTGGGCGGATCACCTCAGGTTAGGATTCAGGACCACTGGGCAATGATGGG 13520
QY 126 AAACCTGTCTCTACTAAATAACAAAATTAGCTGGGCAATGAGGCGATGTCCTGTAT 185
DB 13521 AAACCCCATCTCTACTAAATAACAAAATTAGCTGGGTTGGTGGCACACACCTGTAT 13580
QY 186 CCGAGTACTCGGAGGCTGAGACAGAGAAATTGCTTGAACCTTGGAGGGGAGGTGCA 245
DB 13581 CCGAGTACTTGGAGGCTGGGCAAGAGAAATTAGCTTGAACCTTGGAGGATGGAGGTGCA 13640

RESULT 13
ABAI6094/c
ID ABAI6094 standard; DNA; 19616 BP.
XX

AC ABA16094;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 8425.
XX
KW Human; neutropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antineumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN W0200159063-A2.
XX
PD 16-AUG-2001;
XX
PF 17-JAN-2001; 2001WG-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216920.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233563.
PR 14-SEP-2000; 2000US-0233564.
PR 14-SEP-2000; 2000US-0233565.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241805.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.

```
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI. INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
XX useful for preventing, diagnosing and/or treating nervous system
XX cancers and metastases.
XX
XX Disclosure, SEQ ID NO 8425; 1701pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins
XX (AB14678-AB19001) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (ant)agonists are useful
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pot_sequences.
XX
XX Sequence 19616 BP; 4181 A; 5695 C; 5539 G; 4201 T; 0 other;
XX
XX
XX Query Match 18.8%; Score 188.6; DB 22; Length 19616;
XX Best Local Similarity 96.0%; Pred. No. 3.4e-25;
XX Matches 209; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
XX
XX 3 TATAAAATGCATACCTTTAGGCTGGGCAAGTGGCTGACGCTGTCTGTATATCCAGACTTTG 62
XX 17028 TGTGTAAATGCTAAATGGGCGCCAGGCGAGTGGCTGACGCTGTATATCCAGACTTTG 16969
XX
XX 63 GGAGGCCAAGCTGGGCGGATCACCTGAGGTTAGGATTCAGGACCAACCTGGGCGATCATG 122
XX 16968 GGAGGCCAAGCTGGGCGGATCACCTGAGGTTAGGATTCAGGACCAACCTGGGCGATCATG 16909
XX
XX 123 GCGNAACCTGTCTCTACTAATAAATACAAATAATAGTGGGATGGGAGGCGATGCGCTGT 182
XX 16908 ACGAAACCTGTCTCTACTAATAAATACAAATAATAGTGGGATGGGAGGCGATGCGCTGT 16849
XX
XX 183 AATCCAGCTACTCGGGAGGCTGAGACAGAAGAAATTCCTTGAAGCTTGAGGAGGAGGTT 242
XX 16848 AATCCAGCTACTCGGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGTT 16789
XX
XX 243 GCA 245
XX 16788 GCA 16786
XX
XX RESULT 14
XX ABA16856/c
```

```
ABAI6856 standard; DNA; 19616 BP.
ABAI6856;
23-JAN-2002 (first entry)
Human nervous system related polynucleotide SEQ ID NO 9187.
Human; isotropic; neuroprotective; cytostatic; dermatologic; virucide;
immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
antiparkinsonian; antischling; antianaemic; antiarthritis; cancer;
antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
antiparasitic; cardiant; immune disorder; cardiovascular disorder;
neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
Homo sapiens.
WO200159053-A2.
16-AUG-2001.
17-JAN-2001; 2001WO-US01334.
31-JAN-2000; 2000US-0179065.
04-FEB-2000; 2000US-0180628.
24-FEB-2000; 2000US-0184664.
02-MAR-2000; 2000US-0186350.
16-MAR-2000; 2000US-0189874.
17-MAR-2000; 2000US-0190076.
18-APR-2000; 2000US-0198123.
19-MAY-2000; 2000US-0205515.
07-JUN-2000; 2000US-0209467.
28-JUN-2000; 2000US-0214886.
30-JUN-2000; 2000US-0215135.
07-JUL-2000; 2000US-0216647.
07-JUL-2000; 2000US-0216880.
11-JUL-2000; 2000US-0217487.
11-JUL-2000; 2000US-0217496.
14-JUL-2000; 2000US-0218290.
26-JUL-2000; 2000US-0220963.
26-JUL-2000; 2000US-0220964.
14-AUG-2000; 2000US-0224518.
14-AUG-2000; 2000US-0224519.
14-AUG-2000; 2000US-0225213.
14-AUG-2000; 2000US-0225214.
14-AUG-2000; 2000US-0225266.
14-AUG-2000; 2000US-0225267.
14-AUG-2000; 2000US-0225268.
14-AUG-2000; 2000US-0225270.
14-AUG-2000; 2000US-0225447.
14-AUG-2000; 2000US-0225757.
14-AUG-2000; 2000US-0225758.
14-AUG-2000; 2000US-0225759.
18-AUG-2000; 2000US-0226279.
22-AUG-2000; 2000US-0226681.
22-AUG-2000; 2000US-0226868.
23-AUG-2000; 2000US-0227182.
23-AUG-2000; 2000US-0227009.
30-AUG-2000; 2000US-0228924.
01-SEP-2000; 2000US-0229287.
01-SEP-2000; 2000US-0229343.
01-SEP-2000; 2000US-0229344.
01-SEP-2000; 2000US-0229345.
05-SEP-2000; 2000US-0229509.
05-SEP-2000; 2000US-0229513.
06-SEP-2000; 2000US-0230437.
06-SEP-2000; 2000US-0230438.
08-SEP-2000; 2000US-0231242.
08-SEP-2000; 2000US-0231243.
08-SEP-2000; 2000US-0231244.
08-SEP-2000; 2000US-0231413.
08-SEP-2000; 2000US-0231414.
```

PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246539.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251150.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
RA (HJMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-541565/60.
XX
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,
XX useful for preventing, diagnosing and/or treating nervous system
XX cancers and metastases.
XX
XX Disclosure: SEQ ID NO 9187; 1701pp - Sequence Listing; English.
XX
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins
XX (ABR14578-ABR15001) useful for preventing, treating or ameliorating
XX medical conditions e.g. by protein or gene therapy. The genes are
XX isolated from a range of human tissues disclosed in the specification.
XX The nucleic acids, proteins, antibodies and (antagonists are useful
XX in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias;
XX (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX Note: The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pat_sequences.
XX
XX Sequence 19616 BP; 4181 A; 5695 C; 5539 G; 4201 T; 0 other;
XX
XX Query Match 18.8%; Score 188.6; DB 22; Length 19616;
XX Best Local Similarity 86.0%; Pred. No. 3.4e-25;
XX Matches 209; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
XX
QY 3 TATAAAATGCATACCTTTAGGCTGGGCAAGTGGCTACGTCGTGTATATCCAGCATTG 62
DB 17028 TGTGTAAATGCTAAATGGGCCCAAGCGCAGTGGCTACGCTGTATATCCAGCATTG 16969
QY 63 GGAGGCCCAAGTGGGCGGATCACCTGAGGTAGGATTCAGGACCACTGGCCATCATG 122
DB 16968 GGAGGCCCAAGTGGGCGGATCACCTGAGGTAGGATTCAGGACCACTGGCCATCATG 16909
QY 123 GCGAAACCCCTGCTCTACTAAAAATACAAAAATTAGCTGGGCGATGGAGGCATGTGCCTGT 182
DB 16906 ACGAAACCCCTGCTCTACTAAAAATACAAAAATTAGCTGGGCGATGGAGGCATGTGCCTGT 16849
QY 183 AATCCAGCTACTCGGGAGGCTGAGACAGAGAATTTGCTTGAACCTTGAGGGGGAGGTT 242
DB 16843 AATCCAGCTACTCGGGAGGCTGAGGAGGAGAGATCGCTTGAACCTTGAGGGGGAGGTT 16789
QY 243 GCA 245
DB 16788 GCA 16786

RESULT 15
AAL36855/c
ID AAL36855 standard; DNA; 19616 BP.
XX
AC AAL36855;
XX
DT 09-JAN-2002 (first entry)
XX
DE Human musculoskeletal system related polynucleotide SEQ ID NO 3220.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW anti-allergic; hepatotropic; antidiabetic; anti-inflammatory; antitumor;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiact; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ds.
XX
OS Homo sapiens.
XX
PN WO200153167-A1.
XX
PC 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01338.
XX
PR 31-CAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.

PR 09-SEP-2000; 2000US-0231414.
PR 09-SEP-2000; 2000US-0232060.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235835.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236602.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 09-NOV-2000; 2000US-0246474.
PR 09-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.

```
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-451937/48.
XX
XX Isolated polypeptide for treating, preventing and/or prognosing
XX disorders related to the musculoskeletal system including
XX musculoskeletal cancers and also for testing and detection e.g.
XX diagnosis -
XX
PS Example 2; SEQ ID NO 3220; 761pp + Sequence Listing; English.
XX
CC The invention relates to novel genes (AAL34663-AAL37666' and proteins
CC (AB03087-AB04109) associated with the musculoskeletal system useful
CC for preventing, treating or ameliorating medical conditions e.g. by
CC protein or gene therapy. The genes are isolated from a range of human
CC tissues disclosed in the specification. The nucleic acids, proteins,
CC antibodies and (ant)agonists are useful in the diagnosis, treatment
CC and prevention of: (a) cancer, e.g. breast and ovarian cancer and
CC other cancers of the adrenal gland, bone, bone marrow, breast,
CC gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis;
CC (c) cardiovascular disorders such as myocardial ischaemias; (d) wound
CC healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy;
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 19616 BP; 4181 A; 5695 C; 5339 G; 4201 T; 0 other;

Query Match      18.8%; Score 186.6; DB 22; Length 19616;
Best Local Similarity 86.0%; Pred. No. 3.4e-25;
Matches 209; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 3 TATAAAATGCATACCTTAGGCTGGGCAAGTGGCTCAGCTCTGTATCCAGCACTTGG 62
DB 17028 TGTGTAAATGTAAATGGGCGCAAGCGCAGTGGCTCAGCTCTGTATCCAGCACTTGG 16969

QY 63 GGAGGCCCAAGTGGGCGGATCACCTGAGCTAGGAGTTCAGGACCACTGGCCATCATG 122
DB 16969 GGAGGCCCAAGTGGGCGGATCACCTGAGCTCAGGAGTTCAGGACCACTGGCCACATG 16909

QY 123 GCGAAACCCCTGCTCTACTAAAAATACAAAATTAGCTGGGATGGAGGCACTGTGCTGT 182
DB 16908 ACGAAACCCCTGCTCTACTAAAAATACAAAATTAGCTGGGATGGGACACATGCTGT 16849

QY 183 AATCCAGCTACTCGGAGGCTGAGACAGAAATTGTTGAACCTGGAGGGGAGGTT 242
DB 16848 AATCCAGCTACTCGGAGGCTGAGGAGGAGATCGCTTGAACCTGGAGGGAGGTT 16789

QY 243 GCA 245
|||
```

Db 16788 GCA 16786

Search completed: October 24, 2003, 18:47:25
Job time : 255.249 secs


```

; NAME/KEY: exon 2
; LOCATION: 2115..2262
; FEATURE:
; NAME/KEY: exon 3
; LOCATION: 3056..3202
; FEATURE:
; NAME/KEY: exon 4
; LOCATION: 3387..3472
; FEATURE:
; NAME/KEY: exon 5
; LOCATION: 5667..5923
; FEATURE:
; NAME/KEY: exon 6
; LOCATION: 7745..8955
US-09-077-354B-3

Query Match      19.0%; Score 189.8; DB 3; Length 10380;
Best Local Similarity 86.7%; Pred. No. 1.5e-31;
Matches 209; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 5 TAAAAATGCATACCTTTAGGCTGGGCAAAAGTGGCTCACGTCGTGTGTAATCCCAAGCACTTTGGG 64
   |||||
Db 3540 TAAAAACTTAAGCTCTGCGCGGGCGCAGTGGCTCACGCCGTGTAATCCCAAGCACTTTGGG 3539
   |||||

QY 65 AGGCCAAGGTGGGGGATCACCTGAGGTAGGTTAGGATTCAGGACCAACCTGGCCATCATGGC 124
   |||||
Db 3600 AGGCCAGTTGGGGGATCACCTGAGGTAGGTTAGGATTCAGGACCAACCTGGCCATCATGGT 3659
   |||||

QY 125 GAAACCCCTGCTCTACTAAATAACAAAATTAGCTGGGCATGGAGGATGTGCTGTAA 184
   |||||
Db 3660 GAAACCCCTGCTCTACTAAATAATATAAAATTAGCTGGGCATGGTGGCAGCGGCTGTAA 3719
   |||||

QY 185 TCCAGCTACTCGGGAGGCTGAGACAGAGAGATTGCTTGACCTTGGAGGGGAGGTGC 244
   |||||
Db 3720 TCCCATCTACTAGGAGGCTGAGACAGGAGGATTGCTTAAACCTGGGAGGAGGCTTGC 3779
   |||||

QY 245 A 245
Db 3780 A 3780
```

```

RESULT 2
US-09-791-211-10
; Sequence 10, Application US/09791211
; Patent No. 6448080
; GENERAL INFORMATION:
; APPLICANT: Donna T. Ward
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF WRN EXPRESSION
; FILE REFERENCE: RTS-9205
; CURRENT APPLICATION NUMBER: US/09/791,211
; CURRENT FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 90
; SEQ ID NO 10
; LENGTH: 98844
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 24962
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 64383
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 65468
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 65469
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 65470
; OTHER INFORMATION: unknown
```

```

; NAME/KEY: unsure
; LOCATION: 65471
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 87130
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 89049
; OTHER INFORMATION: unknown
; OTHER INFORMATION:
US-09-791-211-10

Query Match      18.9%; Score 188.8; DB 4; Length 98844;
Best Local Similarity 86.7%; Pred. No. 3.8e-31;
Matches 206; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 6 AAAAAATGCATACCTTTAGGCTGGGCAAAAGTGGCTCACGTCGTGTGTAATCCCAAGCACTTTGGA 65
   |||||
Db 82007 AAAAAATAGCAAAATTTGGCGGGCATGGTGGCTCACGCCGTGTAGTCCCAAGCACTTTGGA 82066
   |||||

QY 66 GGCCAAGGTGGGGGATCACCTGAGGTAGGATTCAGGACCAACCTGGCCATCATGGG 125
   |||||
Db 82067 GGCCGAGGTGGGTGATCACCTGAGGTAGGATTCAGGACCAACCTGGTCAACATGGTG 82126
   |||||

QY 126 AAACCCCTGCTCTACTAAATAACAAAATTAGCTGGGCATGGAGGATGTGCTGTAA 185
   |||||
Db 82127 AAACCCCTGCTCTACTAAATAACAAAATTAGCTGGGCATGGTGGCAGGTGCTGTAA 82186
   |||||

QY 186 CCCAGCTACTCGGGAGGCTGAGACAGAGAGATTGCTTGAACTTGGAGGGGAGGTGGA 245
   |||||
Db 82187 CCCAGCTACTCGGGAGGCTGAGGCGAGGGGAATTGCTTGAACCTCGGAGGTGGAGGTGGA 82246
   |||||

RESULT 3
US-09-800-960-3/c
; Sequence 3, Application US/09800960
; Patent No. 6387677
; GENERAL INFORMATION:
; APPLICANT: YE, Jane et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001158
; CURRENT APPLICATION NUMBER: US/09/800,960
; CURRENT FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 62804
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(62804)
; OTHER INFORMATION: n = A,T,C or G
US-09-800-960-3
```

```

Query Match      18.5%; Score 185.6; DB 4; Length 62804;
Best Local Similarity 89.3%; Pred. No. 2.7e-30;
Matches 200; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 22 GGCTGGGCAAGTGGCTCACCTGTGTAATCCCAAGCACTTTGGAGGGCCCAAGGTGGGGGA 81
   |||||
Db 48950 GGGTGGGCACCGTGGCTCACACCTGTAACTTCAGCACTTTGGGAGGGCCAGCGGGTGA 48891
   |||||

QY 82 TCACCTGAGGTAGGATTCAGGACCAACCTGGCCATCATGGGAAACCCCTGTCTCTACT 141
   |||||
Db 48890 TCACCTGAGGTACAGGATTCAGACCCAGCCCTGGCCACATGTGTGAAACCCCGTCTCTACT 48831
   |||||

QY 142 AAAAAATACAAAATTTAGCTGGGCATGGAGGATGTGCTGTAAATCCCAAGTACTCTGGGAG 201
   |||||
Db 48830 AAAAAATACAAAATTTAGCCGGCATGGTGGCACATGCTGTAAATCCCAAGTACTCTGGGAG 48771
   |||||
```

QY 202 GCTGACACAGAGATTGCTTGAACCTTGGAGGGGAGGTTGCA 245
|||||
Db 48770 GCTGACACAGAGATTGCTTGAACCTTGGAGGGGAGGTTGCA 48727

RESULT 4

US-08-579-445-26
Sequence 26, Application US/08579445
Patent No. 656053
GENERAL INFORMATION:
APPLICANT: Perucho, Manuel
APPLICANT: Peinado, Miguel A.
APPLICANT: Iorov, Yuri
APPLICANT: Malkhasyan, Sergei
TITLE OF INVENTION: Identification of Neoplasms by Detection
of Genetic Deletions
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/579,445
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: US/08/152,484
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Kirkpatrick, Anita M.
REGISTRATION NUMBER: 32,617
REFERENCE/DOCKET NUMBER: STRATAG.009A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 283 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-579-445-26

Query Match 18.4%; Score 184; DB 4; Length 283;
Best Local Similarity 88.8%; Pred. No. 1.3e-30;
Matches 199; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 22 GGCTGGCAACAGTGGCTCACCTGTGTAATCCAGGCACTTTGGAGGGCCAGGTTGGCGGA 81
|||||
Db 1 GGCTGGCGTGGTGGCTCACACTGTATCCAGCACTTTGGAGGGCCAGGTTGGTGGGA 60
|||||
QY 82 TCACCTGAGGTAGGAGTTCAGGACCAACCTGGCCATCATGGGAAACCCCTGTCTACT 141
|||||
Db 61 TCACCTGAGGTAGGAGTTCAGGACCAACCTGGCCATCATGGGAAACCCCTGTCTACT 120
|||||
QY 142 AAAAATACAAAATTAGCTGGGCAATGGAGGCAATGTCCCTGTATCCAGGCTACTCGGGAG 201
|||||
Db 121 AAAAATACAAAATTAGCTGGGCGTGGTGGCGCGCTGTATCCAGGCTACTCGGGAG 180
|||||
QY 202 GCTGACACAGAGAAATTGCTTGAACCTTGGAGGGGAGGTTGCA 245
|||||
Db 181 GCTGAGGCAGGAGAAATCGCTTGAACCCCGGAGGTGGAGGTTGCA 224
|||||

RESULT 5

US-08-133-629-8
Sequence 8, Application US/08133629
Patent No. 5597694
GENERAL INFORMATION:
APPLICANT: Murrece, David J.
APPLICANT: Housman, David E.
TITLE OF INVENTION: AMPLIFICATION OF NUCLEIC ACIDS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
STREET: 600 Atlantic Avenue
CITY: Boston
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 02210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/133,629
FILING DATE: 07-OCT-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Greer, Helen
REGISTRATION NUMBER: 36,816
REFERENCE/DOCKET NUMBER: M0828/7001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-720-3500
TELEFAX: 617-720-2441
TELEX: 92-1742 EZEKIEL
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-133-629-8

Query Match 19.3%; Score 183.6; DB 1; Length 282;
Best Local Similarity 88.4%; Pred. No. 1.6e-30;
Matches 198; Conservative 1; Mismatches 25; Indels 0; Gaps 0;
QY 37 GGCTGGGCAAGTGGCTCACCTGTGTAATCCAGGCACTTTGGAGGGCCAGGTTGGCGGA 81
|||||
Db 1 GGCTGGGCGTGGTGGCTCACACTGTATCCAGCACTTTGGAGGGCCAGGTTGGTGGGA 60
|||||
QY 82 TCACCTGAGGTAGGAGTTCAGGACCAACCTGGCCATCATGGGAAACCCCTGTCTACT 141
|||||
Db 61 TCACCTGAGGTAGGAGTTCAGGACCAACCTGGCCATCATGGGAAACCCCTGTCTACT 120
|||||
QY 142 AAAAATACAAAATTAGCTGGGCAATGGAGGCAATGTCCCTGTATCCAGGCTACTCGGGAG 201
|||||
Db 121 AAAAATACAAAATTAGCTGGGCGTGGTGGCGCGCTGTATCCAGGCTACTCGGGAG 180
|||||
QY 202 GCTGACACAGAGAAATTGCTTGAACCTTGGAGGGGAGGTTGCA 245
|||||
Db 181 GCTGAGGCAGGAGAAATCGCTTGAACCCCGGAGGTGGAGGTTGCA 224
|||||

RESULT 6

US-09-539-333D-37
Sequence 37, Application US/09539333D
Patent No. 6476208
GENERAL INFORMATION:
APPLICANT: Cohen, Daniel
APPLICANT: Blumenfeld, Marta
APPLICANT: Chumakov, Ilya
APPLICANT: Bougueleret, Jydie

```

; APPLICANT: Bihain, Bernard
; APPLICANT: Essioux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: GENSET 047AUS
; CURRENT APPLICATION NUMBER: US/09/539,333D
; CURRENT FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: US 09/416,384
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patent.pm
; SEQ ID NO 37
; LENGTH: 1154
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..719
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 720..1118
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 1119..1154
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 1131..1136
; FEATURE:
; NAME/KEY: allele
; LOCATION: 191
; OTHER INFORMATION: 8-121-197 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 313
; OTHER INFORMATION: 8-122-271 : deletion of CAAA
; FEATURE:
; NAME/KEY: allele
; LOCATION: 314
; OTHER INFORMATION: 8-122-272 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 368
; OTHER INFORMATION: 8-122-326 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 390
; OTHER INFORMATION: 8-123-55 : polymorphic base A or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 814
; OTHER INFORMATION: 8-128-61 : polymorphic base G or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 821
; OTHER INFORMATION: 8-128-63 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 822
; OTHER INFORMATION: 8-128-69 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 838
; OTHER INFORMATION: 8-128-85 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 897
; OTHER INFORMATION: 8-129-50 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 908
; OTHER INFORMATION: 8-129-60 : deletion of A
; US-09-539-333D-37
;
; Query Match
; Best Local Similarity 85.7%; Score 183.6; DB 4; Length 1154;
; Matches 204; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
;
; QY 9 AAATGCATACCTTTAGGCTGGGCAAGTGGCTCAGCTCTGTGTAATCCAGCACCTTTGGGAGG 67
; DB 503 AAGAGGAAGGTCTTGGCCAGGCTCAGTGGCTCAGACCTGTAAATCCAGCACCTTTGGGAGG 562
; QY 68 CCAAGGTGGGCGGATCAGCTGAGGTAGGAGTTAGGAGCAACCTGGCCATCATGGCGAA 127
; DB 563 CCGAGCGGGCGAGATCAGCTGAGGTGAGGAGTTCAAGACCACTGACCAATATGGTGA 622
; QY 128 ACCCTGTCTCTACTAAAAATACAAAATTTAGCTGGGCAATGGAGCATGTGCTGTAATCC 187
; DB 623 ACCCATCTCTACTAAAAATACAAAATTTAGCCAGGCAATGGTGGAGGTGCTGTAGTCC 682
; QY 188 CAGTACTCGGGAGGCTGAGACAGAGAAATTTGTTGAACCTTGGAGGGGGAGGTGCA 245
; DB 583 CAGTACTCGGGAGGCTGAGACAGAGAAATTTGTTGAACCTTGGAGGGGGAGGTGCA 740
;
; RESULT 7
; US-09-539-333D-36
; Sequence 36, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essioux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: GENSET 047AUS
; CURRENT APPLICATION NUMBER: US/09/539,333D
; CURRENT FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: US 09/416,384
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patent.pm
; SEQ ID NO 36
; LENGTH: 1301

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..899
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 900..1265
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 1266..1351
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 1277..1282
; FEATURE:
; NAME/KEY: allele
; LOCATION: 191
; OTHER INFORMATION: 8-121-187 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 313
; OTHER INFORMATION: 8-122-271 : deletion of CAAA
; FEATURE:
; NAME/KEY: allele
; LOCATION: 314
; OTHER INFORMATION: 8-122-272 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 368
; OTHER INFORMATION: 8-122-325 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 390
; OTHER INFORMATION: 8-123-55 : polymorphic base A or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 806
; OTHER INFORMATION: 8-127-28 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 897
; OTHER INFORMATION: 8-127-119 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 937
; OTHER INFORMATION: 8-127-159 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 961
; OTHER INFORMATION: 8-128-61 : polymorphic base G or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 968
; OTHER INFORMATION: 8-128-68 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 969
; OTHER INFORMATION: 8-128-69 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 985
; OTHER INFORMATION: 8-128-85 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1044
; OTHER INFORMATION: 8-129-50 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1055
; OTHER INFORMATION: 8-129-60 : deletion of A
; US-09-539-333D-36
; Query Match 18.3% Score 183.6; DB 4; Length 1301;

Best Local Similarity 85.7%; Pred. No. 2.le-30;
Matches 234; Conservative 3; Mismatches 34; Indels 3; Gaps 3;

QY 8 AAATGCATACCTTTAGGCTGGCCAAAGTGGCTCACCTCTGTAAATCCACGACACTTTGGGAGG 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 503 AAGAGGAAGGTCTTGGCCAGGCTCAGTGGCTCACACCTGTAAATCCGACACTTTGGGAGG 562
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 68 CCAAGGTGGGGGATCACCTGAGGTAGGAGTTGAGGACCAACCTGCGCATCATGCGGAA 127
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 563 CCGAGGCGGGCAGATCACCTGAGGTGAGGATTCAGACACCGCTGACCAATATGTTGAA 622
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 128 ACCCTGTCTTACTAAATAACAAAAATTAGCTGGGATGAGGCAATGCTGCTGTATCC 187
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 623 ACCCATCTCTTACTAAATAACAAAAATTAGCCAGGATGTCGCGAGTGGTGGAGTGGCTGTATCC 682
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 188 CAGCTACTCGGGAGGCTGAGACAGAGAATTGCTTGACCTTGGAGGGGAGGTTGCA 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 643 CAGCTACTCGGGAGGCTGAGACAGAGAATTGCTTGACCTTGGAGGGGAGGTTGCA 749
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
US-09-539-333D-40
; Sequence 40, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essieux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND RIBOSOMAL MARKERS
; FILE REFERENCE: GENSET.047AUS
; CURRENT APPLICATION NUMBER: US/09/539,333D
; CURRENT FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/146,452
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: US 09/416,164
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patent.pr
; SEQ ID NO 40
; LENGTH: 1386
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..984
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 985..1350
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 1351..1386
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 1362..1367
; FEATURE:
; NAME/KEY: allele
```

LOCATION: 191
OTHER INFORMATION: 8-121-187 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 398
OTHER INFORMATION: 8-122-271 : deletion of CAAA
FEATURE:
NAME/KEY: allele
LOCATION: 399
OTHER INFORMATION: 8-122-272 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 453
OTHER INFORMATION: 8-122-326 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 475
OTHER INFORMATION: 8-123-55 : polymorphic base A or T
FEATURE:
NAME/KEY: allele
LOCATION: 891
OTHER INFORMATION: 8-127-28 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 982
OTHER INFORMATION: 8-127-119 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 1022
OTHER INFORMATION: 8-127-159 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 1046
OTHER INFORMATION: 8-128-61 : polymorphic base G or C
FEATURE:
NAME/KEY: allele
LOCATION: 1053
OTHER INFORMATION: 8-128-69 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 1054
OTHER INFORMATION: 8-128-69 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 1070
OTHER INFORMATION: 8-128-85 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 1129
OTHER INFORMATION: 8-129-50 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 1140
OTHER INFORMATION: 9-129-60 : deletion of A
US-09-539-333D-4C

Query Match: 18.3%; Score 183.6; DB 4, Length 1386;
Best Local Similarity 85.7%; Pred. No. 2, e-30;
Matches 204; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
QY 5 AATGTCATCTTAGGCTGGGCAAGTGGCTCAGCTCTGTATCCAGCAGCTTTGGGAGG 67
DB 588 AAGAGGAAGTCTTGGCCAGGCTCAGTGGCTCACACCTGTATCCAGCAGCTTTGGGAGG 647
QY 68 CCAAGGTGGCGGATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAA 127
DB 648 CCGAGCGGGCAGATCACCTGAGTTCAGGAGTCAAGACCAACCTGACCATATGCTGAA 707
QY 128 ACCGTGCTCTACTAAAATACAAAATTAGCTGGGCATGAGGCGATGTGCTGTATCC 187
DB 708 ACCCATCTCTACTAAAATACAAAATTAGCAGGCGATGTTGGCAGGTCCTGTATCC 767
QY 188 CAGCTACTCGGGAGGCTGAGACAGGAAGTTCCTTGAACCTTGGAGGGGGAGGTTGCA 245

DB 768 CAGCTACTCGGGAGGCTGAGACAGGAAGTTCCTTGAACCTTGGAGGGGGAGGTTGCA 825
RESULT 9
US-09-679-409-1
Sequence 1, Application US/09679409
Patent No. 6555316
GENERAL INFORMATION:
APPLICANT: Cohen, Daniel
APPLICANT: Blumenfeld, Marta
APPLICANT: Chumakov, Ilya
APPLICANT: Bougueleret, Lydie
APPLICANT: Essioux, Laurent
TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENE, PROTEINS AND BIALLELIC MARKERS
FILE REFERENCE: 53-US15.CIP
CURRENT APPLICATION NUMBER: US/09/679,409
CURRENT FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 09/539,333
PRIOR FILING DATE: 2000-03-03
PRIOR APPLICATION NUMBER: 09/416,384
PRIOR FILING DATE: 1999-10-12
PRIOR APPLICATION NUMBER: 60/168,088
PRIOR FILING DATE: 1999-11-30
NUMBER OF SEQ ID NOS: 134
SOFTWARE: Patent.pm
SEQ ID NO 1
LENGTH: 319608
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 19122..201122
OTHER INFORMATION: 5' regulatory region
NAME/KEY: exon
LOCATION: 201123..201234
OTHER INFORMATION: exon S
NAME/KEY: exon
LOCATION: 201123..201560
OTHER INFORMATION: exon S2
NAME/KEY: exon
LOCATION: 214676..214793
OTHER INFORMATION: exon T
NAME/KEY: exon
LOCATION: 215702..215746
OTHER INFORMATION: exon U
NAME/KEY: exon
LOCATION: 216836..216994
OTHER INFORMATION: exon V
NAME/KEY: exon
LOCATION: 216836..217077
OTHER INFORMATION: exon V2
NAME/KEY: exon
LOCATION: 217671..217764
OTHER INFORMATION: exon V1
NAME/KEY: exon
LOCATION: 227655..227736
OTHER INFORMATION: exon V4
NAME/KEY: exon
LOCATION: 239715..238919
OTHER INFORMATION: exon V3
NAME/KEY: exon
LOCATION: 240440..240673
OTHER INFORMATION: exon W
NAME/KEY: exon
LOCATION: 240440..241153
OTHER INFORMATION: exon W2
NAME/KEY: exon
LOCATION: 241072..241291
OTHER INFORMATION: exon X
NAME/KEY: exon
LOCATION: 244353..244561
OTHER INFORMATION: exon Y


```

1 LOCATION: 231797..231880
2 OTHER INFORMATION: exon 02 complement g34872 gene
3 FEATURE:
4 NAME/KEY: exon
5 LOCATION: 231870..231879
6 OTHER INFORMATION: exon 01 complement g34872 gene
7 FEATURE:
8 NAME/KEY: exon
9 LOCATION: 234174..234321
10 OTHER INFORMATION: exon 0 complement g34872 gene
11 FEATURE:
12 NAME/KEY: exon
13 LOCATION: 237406..237428
14 OTHER INFORMATION: exon Nbis complement g34872 gene
15 FEATURE:
16 NAME/KEY: exon
17 LOCATION: 239719..239807
18 OTHER INFORMATION: exon N2 complement g34872 gene
19 FEATURE:
20 NAME/KEY: exon
21 LOCATION: 239719..239853
22 OTHER INFORMATION: exon N complement g34872 gene
23 FEATURE:
24 NAME/KEY: exon
25 LOCATION: 243528..243569
26 OTHER INFORMATION: exon N117 complement g34872 gene
27 FEATURE:
28 NAME/KEY: exon
29 LOCATION: 240528..240617
30 OTHER INFORMATION: exon M1069 complement g34872 gene
31 FEATURE:
32 NAME/KEY: exon
33 LOCATION: 240528..240644
34 OTHER INFORMATION: exon M52 complement g34872 gene
35 FEATURE:
36 NAME/KEY: exon
37 LOCATION: 240528..240824
38 OTHER INFORMATION: exon M862 complement g34872 gene
39 FEATURE:
40 NAME/KEY: exon
41 LOCATION: 240528..240994
42 OTHER INFORMATION: exon M692 complement g34872 gene
43 FEATURE:
44 NAME/KEY: exon
45 LOCATION: 240528..241685
46 OTHER INFORMATION: exon M1 complement g34872 gene
47 FEATURE:
48 NAME/KEY: exon
49 LOCATION: 243800..243993
50 OTHER INFORMATION: exon M51 complement g34872 gene
51 FEATURE:
52 NAME/KEY: misc feature
53 LOCATION: 241686..243685
54 OTHER INFORMATION: 5' regulatory region g34872 gene
55 FEATURE:
56 NAME/KEY: misc feature
57 LOCATION: 290652..292652
58 OTHER INFORMATION: 3' regulatory region g34665 gene
59 FEATURE:
60 NAME/KEY: exon
61 LOCATION: 292653..292841

```

```

Query Match      19.1%; Score 193.2; DB 4; Length 319638;
Best Local Similarity 85.1%; Pred. No. 7.3e-30;
Matches 203; Conservative 1; Mismatches 34; Indels 0; Gaps 0;

QY      8  AAATGCATACTTTCAGGCTGGGCAAGTGGGCTCAGCTCTGTATCCGACGACCTTGGGAGG 67
Db      29990  AAGAGGAAGGCTCTTGGCAGGCTCAGTGGCTCACCTGTATCCGACGACCTTGGGAGG 30019

```

```

QY      66  CCAGGTGGGGGATCACCTGAGGTTAGAGTTCAGAGACCAACCTGGGCCATCATGCGGAA 127
DB      30040  CCGAGGCGGGGAGATCACCTGAGGTTAGAGTTCAGAGACCAACCTGGGCCATCATGCGGAA 30039
QY      129  ACCCTGTCTCTACTAAAAATACAAAATTAGCTGGGATGGAGGATGTGCTGTAAATCC 187
DB      30100  ACCCCATCTCTACTAAAAATACAAAATTAGCCAGGCATGCTGGAGGTGCTGTAAATCC 30099
QY      188  CAGTACTCGGAGGCTGAGACAGAGAAATTCCTTGAACCTTGGAGGGGAGGTTGCA 245
DB      30160  CAGTACTCGGAGGCTGAGACAGGAGAAATTCCTTGAACCTTGGAGGGGAGGTTGCA 30217

RESULT 1:
US-09-750-580-1/c
: Sequence 1, Application US/09750580
: Patent No. 6455280
: GENERAL INFORMATION:
:   APPLICANT: Yen, Frances
:   APPLICANT: Denison, Blake
:   APPLICANT: Bour, Barbara
:   APPLICANT: Birkain, Bernard
:   APPLICANT: Dumas Milne Edwards, Jean-Baptiste
:   APPLICANT: Duclert, Aymeric
:   APPLICANT: Bougueleret, Lydie
:   APPLICANT: Ebbets-Reed, Dana
:   APPLICANT: Salter-Cid, Luisa
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL GROWTH
: FILE REFERENCE: 89 US2.CIP
: CURRENT APPLICATION NUMBER: US/09/750,580
: CURRENT FILING DATE: 2000-12-28
: PRIOR APPLICATION NUMBER: US 09/599,362
: PRIOR FILING DATE: 2000-06-21
: PRIOR APPLICATION NUMBER: PCT/B00/0101
: PRIOR FILING DATE: 2000-06-21
: PRIOR APPLICATION NUMBER: PCT/IB99/02058
: PRIOR FILING DATE: 1999-12-20
: PRIOR APPLICATION NUMBER: US 49/469/059
: PRIOR FILING DATE: 1999-12-21
: PRIOR APPLICATION NUMBER: US 60/113,686
: PRIOR FILING DATE: 1998-12-22
: PRIOR APPLICATION NUMBER: US 60/141,032
: PRIOR FILING DATE: 1998-06-25
: NUMBER OF SEQ ID NOS: 6
: SOFTWARE: Patent.pm
: SEQ ID NO: 1
: LENGTH: 81001
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
:   NAME/KEY: misc_feature
:   LOCATION: 10946..12946
:   OTHER INFORMATION: 5' regulatory region
:   NAME/KEY: exon
:   LOCATION: 12947..12958
:   OTHER INFORMATION: exon 1
:   NAME/KEY: exon
:   LOCATION: 13470..13526
:   OTHER INFORMATION: exon 2
:   NAME/KEY: exon
:   LOCATION: 13641..13752
:   OTHER INFORMATION: exon 3
:   NAME/KEY: exon
:   LOCATION: 14271..15969
:   OTHER INFORMATION: exon 4
:   NAME/KEY: misc_feature
:   LOCATION: 15969..17969
:   OTHER INFORMATION: 3' regulatory region
:   NAME/KEY: allele
:   LOCATION: 1239
:   OTHER INFORMATION: 20-828-311 : polymorphic base C or T
:   NAME/KEY: allele

```

LOCATION: 12347
OTHER INFORMATION: 17-42-319 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 15241
OTHER INFORMATION: 17-41-250 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 42218
OTHER INFORMATION: 20-841-149 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 45442
OTHER INFORMATION: 20-842-115 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 77058
OTHER INFORMATION: 20-853-415 : polymorphic base C or T
NAME/KEY: primer_bind
LOCATION: 929..949
OTHER INFORMATION: 20-828.pu
NAME/KEY: primer_bind
LOCATION: 1357..1377
OTHER INFORMATION: 20-828.rp complement
NAME/KEY: primer_bind
LOCATION: 12029..12050
OTHER INFORMATION: 17-42.pu
NAME/KEY: primer_bind
LOCATION: 12581..12603
OTHER INFORMATION: 17-42.rp complement
NAME/KEY: primer_bind
LOCATION: 14992..15012
OTHER INFORMATION: 17-41.pu
NAME/KEY: primer_bind
LOCATION: 15460..15482
OTHER INFORMATION: 17-41.rp complement
NAME/KEY: primer_bind
LOCATION: 42070..42090
OTHER INFORMATION: 20-841.pu
NAME/KEY: primer_bind
LOCATION: 42572..42591
OTHER INFORMATION: 20-841.rp complement
NAME/KEY: primer_bind
LOCATION: 45328..45347
OTHER INFORMATION: 20-842.pu
NAME/KEY: primer_bind
LOCATION: 45863..45883
OTHER INFORMATION: 20-842.rp complement
NAME/KEY: primer_bind
LOCATION: 76644..76664
OTHER INFORMATION: 20-853.pu
NAME/KEY: primer_bind
LOCATION: 77166..77185
OTHER INFORMATION: 20-853.rp complement
NAME/KEY: primer_bind
LOCATION: 1220..1229
OTHER INFORMATION: 20-828-311.mis
NAME/KEY: primer_bind
LOCATION: 1240..1258
OTHER INFORMATION: 20-828-311.mis complement
NAME/KEY: primer_bind
LOCATION: 12328..12346
OTHER INFORMATION: 17-42-319.mis
NAME/KEY: primer_bind
LOCATION: 12348..12366
OTHER INFORMATION: 17-42-319.mis complement
NAME/KEY: primer_bind
LOCATION: 15222..15240
OTHER INFORMATION: 17-41-250.mis
NAME/KEY: primer_bind
LOCATION: 15242..15260
OTHER INFORMATION: 17-41-250.mis complement
NAME/KEY: primer_bind
LOCATION: 42199..42217
OTHER INFORMATION: 20-841-149.mis
NAME/KEY: primer_bind
LOCATION: 42219..42237

OTHER INFORMATION: 20-841-149.mis complement
NAME/KEY: primer_bind
LOCATION: 45423..45441
OTHER INFORMATION: 20-842-115.mis
NAME/KEY: primer_bind
LOCATION: 45443..45461
OTHER INFORMATION: 20-842-115.mis complement
NAME/KEY: primer_bind
LOCATION: 77039..77057
OTHER INFORMATION: 20-853-415.mis
NAME/KEY: primer_bind
LOCATION: 77059..77077
OTHER INFORMATION: 20-853-415.mis complement
NAME/KEY: misc binding
LOCATION: 1227..1251
OTHER INFORMATION: 20-828-311.probe
NAME/KEY: misc binding
LOCATION: 12335..12359
OTHER INFORMATION: 17-42-319.probe
NAME/KEY: misc binding
LOCATION: 15229..15253
OTHER INFORMATION: 17-41-250.probe
NAME/KEY: misc binding
LOCATION: 42206..42230
OTHER INFORMATION: 20-841-149.probe
NAME/KEY: misc binding
LOCATION: 45430..45454
OTHER INFORMATION: 20-842-115.probe
NAME/KEY: misc binding
LOCATION: 77046..77070
OTHER INFORMATION: 20-853-415.probe
US-09-750-530-1

Query Match 18.2%; Score 182.4; DB 4; Length 81001;
Best Local Similarity 85.0%; Pred. No. 8.3e-30;
Matches 204; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 6 AAAAATGCATACCTTTAGCGCTGGGCAAGTGGCTCAGTCTGTATATCCAGCACCTTTGGGA 65
DB 65352 AATAAGGAATTTGTGGCCAGGTGCAGTGGCTCAGGCTGTATATCCAGCACCTTTGGGA 64993
QY 66 GGCCAAAGTGGCGGATCACCTGAGGTAGGATTAGGACCAACCTGGGCATCATGGCG 125
DB 64992 GGCCAAAGTGGGTGGATCACCTGAGGTAGGATTAGGACCAACCTGGGCATCATGGCG 64933
QY 126 AAACCCCTGCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCTGTAAAT 185
DB 64932 AAACCCCTGCTCTACTAAAAATACAAAAATTAGCGGGCATGGTGGCAGGAGCCTGTAAAT 64873
QY 186 CCCAGTACTCGGAGGCTGAGACAGAGAGATTTGTAACCTTGGAGGGGGAGGTTGCA 245
DB 64872 CCCAGTACTCAGGAGGCTGAGGAGGAGATCACTTGAACCCCGGAGGGAGGTTGCA 64813

RESULT 12
US-09-499-522-1
Sequence 1, Application US/09499522
Patent No. 6479238
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Martha
APPLICANT: Bougueleret, Lydie
APPLICANT: Sihain, Bernard
TITLE OF INVENTION: POLYMORPHIC MARKERS OF THE LSR GENE
FILE REFERENCE: GENSET.053AUS
CURRENT APPLICATION NUMBER: US/09/499,522
CURRENT FILING DATE: 2000-02-10
EARLIER APPLICATION NUMBER: US 60/119,592
EARLIER FILING DATE: 1999-02-10
EARLIER APPLICATION NUMBER: US 60/144,784
EARLIER FILING DATE: 1999-07-20
NUMBER OF SEQ ID NOS: 20
SOFTWARE: Patent.pm
SEQ ID NO 1

```
1 LENGTH: 23187
2 TYPE: DNA
3 ORGANISM: Homo sapiens
4 FEATURE:
5 NAME/KEY: exon
6 LOCATION: 2001..2356
7 OTHER INFORMATION: exon:
8 FEATURE:
9 NAME/KEY: exon
10 LOCATION: 3543..3884
11 OTHER INFORMATION: exon:2
12 FEATURE:
13 NAME/KEY: exon
14 LOCATION: 12163..12282
15 OTHER INFORMATION: exon:3
16 FEATURE:
17 NAME/KEY: exon
18 LOCATION: 15144..15200
19 OTHER INFORMATION: exon:4
20 FEATURE:
21 NAME/KEY: exon
22 LOCATION: 15765..15911
23 OTHER INFORMATION: exon:5
24 FEATURE:
25 NAME/KEY: exon
26 LOCATION: 19579..19752
27 OTHER INFORMATION: exon:6
28 FEATURE:
29 NAME/KEY: exon
30 LOCATION: 19899..19958
31 OTHER INFORMATION: exon:7
32 FEATURE:
33 NAME/KEY: exon
34 LOCATION: 20056..20187
35 OTHER INFORMATION: exon:8
36 FEATURE:
37 NAME/KEY: exon
38 LOCATION: 20329..20957
39 OTHER INFORMATION: exon:9
40 FEATURE:
41 NAME/KEY: exon
42 LOCATION: 21047..21187
43 OTHER INFORMATION: exon:10
44 FEATURE:
45 NAME/KEY: polyA signal:
46 LOCATION: 21168..21173
47 OTHER INFORMATION: AATAAA
48 FEATURE:
49 NAME/KEY: misc_feature
50 LOCATION: 1..2000
51 OTHER INFORMATION: potential 5' regulatory region
52 FEATURE:
53 NAME/KEY: misc_feature
54 LOCATION: 22324..23187
55 OTHER INFORMATION: homology with USF2 gene in ref: e-b1 Y0756:
56 FEATURE:
57 NAME/KEY: primer_bind
58 LOCATION: 523..544
59 OTHER INFORMATION: upstream amplification primer 17-2
60 FEATURE:
61 NAME/KEY: primer_bind
62 LOCATION: 1047..1068
63 OTHER INFORMATION: downstream amplification primer 17-2 , complement
64 FEATURE:
65 NAME/KEY: primer_bind
66 LOCATION: 346..363
67 OTHER INFORMATION: upstream amplification primer 99-4576
68 FEATURE:
69 NAME/KEY: primer_bind
70 LOCATION: 1385..1402
71 OTHER INFORMATION: downstream amplification primer 99-4576 , complement
72 NAME/KEY: primer_bind
73 LOCATION: 7995..8012
```

```
1 LOCATION: 1096..1115
2 OTHER INFORMATION: upstream amplification primer 9-19
3 FEATURE:
4 NAME/KEY: primer_bind
5 LOCATION: 1616..1635
6 OTHER INFORMATION: downstream amplification primer 9-19 , complement
7 FEATURE:
8 NAME/KEY: primer_bind
9 LOCATION: 1602..1621
10 OTHER INFORMATION: upstream amplification primer 9-20
11 FEATURE:
12 NAME/KEY: primer_bind
13 LOCATION: 2074..2093
14 OTHER INFORMATION: downstream amplification primer 9-20 , complement
15 FEATURE:
16 NAME/KEY: primer_bind
17 LOCATION: 2036..2053
18 OTHER INFORMATION: upstream amplification primer 99-4557
19 FEATURE:
20 NAME/KEY: primer_bind
21 LOCATION: 2563..2580
22 OTHER INFORMATION: downstream amplification primer 99-4557 , complement
23 FEATURE:
24 NAME/KEY: primer_bind
25 LOCATION: 2084..2102
26 OTHER INFORMATION: upstream amplification primer 9-1
27 FEATURE:
28 NAME/KEY: primer_bind
29 LOCATION: 2483..2500
30 OTHER INFORMATION: downstream amplification primer 9-1 , complement
31 FEATURE:
32 NAME/KEY: primer_bind
33 LOCATION: 2470..2489
34 OTHER INFORMATION: upstream amplification primer 9-21 , complement
35 FEATURE:
36 NAME/KEY: primer_bind
37 LOCATION: 2062..2081
38 OTHER INFORMATION: downstream amplification primer 9-21
39 FEATURE:
40 NAME/KEY: primer_bind
41 LOCATION: 3455..3474
42 OTHER INFORMATION: upstream amplification primer 9-3
43 FEATURE:
44 NAME/KEY: primer_bind
45 LOCATION: 3882..3901
46 OTHER INFORMATION: downstream amplification primer 9-3 , complement
47 FEATURE:
48 NAME/KEY: primer_bind
49 LOCATION: 3775..3792
50 OTHER INFORMATION: upstream amplification primer 99-4558
51 FEATURE:
52 NAME/KEY: primer_bind
53 LOCATION: 4336..4356
54 OTHER INFORMATION: downstream amplification primer 99-4558 , complement
55 FEATURE:
56 NAME/KEY: primer_bind
57 LOCATION: 4902..4920
58 OTHER INFORMATION: upstream amplification primer 99-14419 , complement
59 FEATURE:
60 NAME/KEY: primer_bind
61 LOCATION: 4444..4463
62 OTHER INFORMATION: downstream amplification primer 99-14419
63 FEATURE:
64 NAME/KEY: primer_bind
65 LOCATION: 6638..6655
66 OTHER INFORMATION: upstream amplification primer 99-4577
67 FEATURE:
68 NAME/KEY: primer_bind
69 LOCATION: 7072..7089
70 OTHER INFORMATION: downstream amplification primer 99-4577 , complement
71 FEATURE:
72 NAME/KEY: primer_bind
73 LOCATION: 7995..8012
```

```
OTHER INFORMATION: upstream amplification primer 99-4559
FEATURE:
NAME/KEY: primer bind
LOCATION: 8576..8593
OTHER INFORMATION: downstream amplification primer 99-4559, complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 9622..9639
OTHER INFORMATION: upstream amplification primer 99-3148
FEATURE:
NAME/KEY: primer bind
LOCATION: 10023..10040
OTHER INFORMATION: downstream amplification primer 99-3148, complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 9964..9981
OTHER INFORMATION: upstream amplification primer 99-4560
FEATURE:
NAME/KEY: primer bind
LOCATION: 10546..10563
OTHER INFORMATION: downstream amplification primer 99-4560, complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 10996..11015
OTHER INFORMATION: upstream amplification primer 99-14411, complement
FEATURE:
NAME/KEY: primer bind
LOCATION: 10492..10512
OTHER INFORMATION: downstream amplification primer 99-14411
FEATURE:
NAME/KEY: primer bind
LOCATION: 11972..11990
OTHER INFORMATION: upstream amplification primer 99-4561
FEATURE:
NAME/KEY: primer bind
LOCATION: 12481..12501
OTHER INFORMATION: downstream amplification primer 99-4561, complement

Query Match      18.2%  Score 181.8;  DB 4;  Length 23187;
Best Local Similarity  88.0%  Pred. No. 8.8e-30;
Matches 198;  Conservative  0;  Mismatches  27;  Indels  0;  Gaps  0;

QY  21  AGGCTGGGCAAGTGGCTCAGCTCTGTATATCCAGCACTTTGGAGGCCAAGGTGGCGG  80
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  5591 AGTCTGGGACGGTGGCTCAGCTCTGTATATCCAGCACTTTGGAGGCCAAGGTGGG  5650
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY  81  ATCAGCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGTGGGAAACCTGTCTCTAC  140
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  5651 ATCAGCTGAGGTCAGGAGTTCAGGACCAACCTGGCCATCATGTGGGAAACCTGTCTCTAC  5710
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY  141  TAAAAATACAAAATTAGCTGGGCATGGAGGCATGTGGTGTATATCCAGCTACTCGGA  200
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  5711 TAAAAATACAAAATTAGCTGGGCATGGAGGCATGTGGTGTATATCCAGCTACTCGGA  5770
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY  201  GGCTGAGACAGAGAATTCTGTGAACCTTGGAGGGGGAGGTTTSCA  245
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  5771 GGCTGAGGATGAGATCTGTGAACCTTGGAGGGGGAGGTTTSCA  5815
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 13
US-09-820-002-3/c
; Sequence 3, Application US/0982002
; Patent No. 6482630
; GENERAL INFORMATION:
; APPLICANT: Gan, Weiniu
; APPLICANT: Ye, Jane
; APPLICANT: DiFrancesco, Valentina
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001194
; CURRENT APPLICATION NUMBER: US/09/820,002
```

```
CURRENT FILING DATE: 2001-03-29
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 21784
TYPE: DNA
ORGANISM: HUMAN
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(21784)
OTHER INFORMATION: n = A,T,C or G
US-09-820-002-3

Query Match      19.0%  Score 180.6;  DB 4;  Length 21784;
Best Local Similarity  84.0%  Pred. No. 1.6e-29;
Matches 204;  Conservative  0;  Mismatches  39;  Indels  0;  Gaps  0;

QY  3  TATAAAATGCATATCTTAGGCTGGGCAAGTGGCTCAGCTCTGTAATCCAGCACTTG  62
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  18354 TAAAAAAAACCAATAGGCTGGGCAAGTGGCTCAGCTCTGTAATCCAGCACTTG  18295
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY  63  GGAGGCCAAGGTGGCGGATCACCTGAGGTTAGGATTCAGGACCAACCTGGCCATCATG  122
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  18294 GGAGGCCAAGGTGGCGGATCACCTGAGGTTAGGATTCAGGACCAACCTGGCCATCATG  18235
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY  123  GCGAAACCTGTCTCTACTAAAAATACAAAATAGTTCAGGCAATTCGTCGGGCAATGGAGGATGTGCCTGT  182
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  18234 GCGAAACCTGTCTCTACTAAAAATACAAAATAGTTCAGGCAATTCGTCGGGCAATGGAGGATGTGCCTGT  18175
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY  183  AATCCCAGCTACTCGGAGGCTGAGACAGAGAAATTCGTCGGGCAATTCGTCGGGCAATGGAGGATGTGCCTGT  242
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  18174 AATCCCAGCTACTCGGAGGCTGAGACAGAGAAATTCGTCGGGCAATTCGTCGGGCAATGGAGGATGTGCCTGT  18115
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY  243  GCA 245
      |||
Db  18114 GCA 18112
      |||

RESULT 14
US-09-443-184-35
; Sequence 35, Application US/03443184A
; Patent No. 6372431
; GENERAL INFORMATION:
; APPLICANT: Cunningham, Mary Jane
; APPLICANT: Zweiger, Gary
; APPLICANT: Kaser, Matthew R.
; APPLICANT: Panzer, Scott
; APPLICANT: Seilhammer, Jeffrey J.
; APPLICANT: Yue, Henry
; APPLICANT: Baughn, Mariah
; APPLICANT: Azimzai, Yalda
; APPLICANT: Lai, Preeti
; TITLE OF INVENTION: MAXYALIAN TOXICOLOGICAL RESPONSE MARKERS
; FILE REFERENCE: PC-0007 US
; CURRENT APPLICATION NUMBER: US/09/443,184A
; CURRENT FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: PEARL Program
; SEQ ID NO 35
; LENGTH: 1762
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6372431 2742442CB1
US-09-443-184-35

Query Match      19.0%  Score 180.2;  DB 4;  Length 1762;
Best Local Similarity  84.2%  Pred. No. 1.2e-29;
Matches 203;  Conservative  0;  Mismatches  38;  Indels  0;  Gaps  0;

QY  5  TAAAAATGCATATCTTAGGCTGGGCAAGTGGCTCAGCTCTGTAATCCAGCACTTTGG  64
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

Dh 1458 TTAAGTAAAJAACTTTGGGCTGGGCACAGTGCGCTCACACCTGTAAATCCAGGCACTTGGGG 1517
QY 65 AGGCCAAGGTGGCGGATCAGCTGAGGTTAGAGTTTACAGGACCAACCTGACCATCAATGGC 124
Dh 1518 AAGCTGAGGTGGCAGATCAGTTGAGGTGAGGTCAGGAGTTCTAGACCAAGCTGTCAACATGGC 1577
QY 125 GAAACCTGTCTCTACTAAATAACAAATTAAGTGGGATGGAGGCAATGTGCTGTAA 184
Dh 1578 GAAACCCCATCTCTACTAAATAACAAATTAAGTGGGCTGTGGCAGACACACCTGTAA 1637
QY 185 TCCAGCTACTCGGAGGCTGACACAGAGAAATTCCTTGAACTTGGAGGGGAGGTTGC 244
Dh 1638 TCCAGTACTTGGAGGCTGAGGACACAGAAATTCGCTTGAACCCGGGAGGCTTGC 1697
QY 245 A 245
Dh 1698 A 1698

RESULT 15

US-09-491-356C-1/c
Sequence 1, Application US/09491356C
Patent No. 6566061
GENERAL INFORMATION:
APPLICANT: Philibert, Robert A.
APPLICANT: Ginn, Edward I.
APPLICANT: Delist, Lynn
TITLE OF INVENTION: IDENTIFICATION OF POLYMORPHISMS IN THE PCTG4 REGION OF XQ23
FILE REFERENCE: 9465.6US:1
CURRENT APPLICATION NUMBER: US/09/491.356C
CURRENT FILING DATE: 2000-01-26
PRIOR APPLICATION NUMBER: PCT/US99/09365
PRIOR FILING DATE: 1999-04-29
PRIOR APPLICATION NUMBER: 60/083,465
PRIOR FILING DATE: 1998-04-29
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 55298
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (485)..(485)
OTHER INFORMATION: n is not determined
NAME/KEY: misc feature
LOCATION: (838)..(838)
OTHER INFORMATION: n is not determined
NAME/KEY: misc feature
LOCATION: (16728)..(16728)
OTHER INFORMATION: n is not determined
NAME/KEY: misc feature
LOCATION: (22750)..(22750)
OTHER INFORMATION: n is not determined
NAME/KEY: misc feature
LOCATION: (22756)..(22756)
OTHER INFORMATION: n is not determined
NAME/KEY: misc feature
LOCATION: (28519)..(28519)
OTHER INFORMATION: n is not determined
NAME/KEY: misc feature
LOCATION: (44804)..(44804)
OTHER INFORMATION: n is not determined
NAME/KEY: misc feature
LOCATION: (45002)..(45002)
OTHER INFORMATION: n is not determined
NAME/KEY: misc feature
LOCATION: (54049)..(54049)
OTHER INFORMATION: n is not determined
NAME/KEY: misc feature
LOCATION: (54226)..(54226)
OTHER INFORMATION: n is not determined
US-09-491-356C-1

Query Match 18.0%; Score 180; DB 4; Length 55298;
Best Local Similarity 83.6%; Pred. No. 2.5e-29;
Matches 204; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
QY 2 ATATAAAATGCAATCTTAGGCTGGGCAAGTGGCTCAGCTCTGTATATCCAGGCACTTT 61
Dh 43198 ATCTAGAAAAGCACCTTAGAGGCTGGGCGTGATGGCTGATGCCTGTATATCCAGGCACTTT 43139
QY 62 GGGAGGCCAAGGTGGCGGATCACCTGAGGTAGGATTCAGGACCAACCTGGGCAATCAT 121
Dh 43139 GGGAGGCCGAGGAGGTGGATCACCTGAGGTGAGGATTCAGGACCAACCTGGGCAACAT 43079
QY 122 GGGGAAACCCCTGTCTCTACTAAAAAATACAAAAATTAGGTGGGCAATGAGGCAATGTGCTTG 181
Dh 43078 GGTGAAACCCCTGTCTCTACTAAAAAATACAAAAATATGCGGGGTGTGGTGGCAGGCGCTTG 43019
QY 182 TAATCCAGCTACTCGGAGGCTGAGACAGAGAAATTTGTAACCTTGGAGGGGGAGGT 241
Dh 43018 TAATCCAGCTACTCGGAGGCTGAGGAGGCTGAGGCAAGAGATTGCTTTAACTCCGGGAGGCGGAGGT 42959
QY 242 TGCA 245
Dh 42958 TGTA 42955

Search completed: October 24, 2003, 15:43:01
Job time : 66.6458 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

CM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:28:54 ; Search time 246.585 seconds
(without alignments)
10886.662 Million cell updates/sec

File: US-09-830-902-1_COPY_100000_101000

Perfect score: 1001

Sequence: : aatataaaaatgcatactctt.....acctgataatgggtctctctttt 1001

Scoring table: IDENTITY_NUC

Gapop 10.0 ; Gapext 1.0

Searched: 1792395 seqs, 134090451 residues

Total number of hits satisfying chosen parameters: 1384790

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA *

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PTC_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PTCUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB3.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB3.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB3.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10P_PUBCOMB3.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB3.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	191.4	19.1	3143	22	US-10-275-910-1 Sequence 1, Appl:
2	191.4	19.1	30573	11	US-09-764-891-6869 Sequence 6869, Ap
3	191.4	19.1	32195	11	US-09-764-891-6869 Sequence 6868, Ap
4	190.8	19.1	7677	12	US-09-814-353-21993 Sequence 21993, A
5	189.8	19.0	10380	11	US-09-836-613-3 Sequence 3, Appl:
6	189.4	18.9	1325	13	US-10-027-632-123772 Sequence 123772,
7	189.4	18.9	1325	13	US-10-027-632-123772 Sequence 123773,
8	189.4	18.9	2490	13	US-10-027-632-102144 Sequence 102144,
9	189.4	18.9	2490	13	US-10-027-632-111607 Sequence 111607,
10	188.8	18.9	614	13	US-10-027-632-306904 Sequence 306904,
11	188.8	18.9	614	13	US-10-027-632-306905 Sequence 306905,
12	188.8	18.9	944	13	US-10-027-632-77917 Sequence 77917, A
13	188.8	18.9	17792	9	US-09-764-869-1599 Sequence 1599, Ap
14	188.8	18.9	17792	14	US-10-091-504-1599 Sequence 1599, Ap
15	188.8	18.8	17792	14	US-09-764-877-3220 Sequence 3220, Ap
16	188.6	18.8	19616	10	US-09-764-877-3220 Sequence 3220, Ap

C	17	187.6	18.7	11754	11	US-09-984-827-5	Sequence 5, Appl:
	18	186.6	18.6	17335	10	US-09-764-847-1280	Sequence 1280, Ap
	19	186.6	18.6	17335	14	US-10-092-154-1280	Sequence 1280, Ap
	20	186.6	18.6	19892	10	US-09-764-847-1281	Sequence 1281, Ap
	21	186.6	18.6	19882	14	US-10-092-154-1281	Sequence 1281, Ap
	22	186.6	18.6	116840	14	US-10-020-141-3	Sequence 3, Appl:
	23	186.6	18.6	1691139	14	US-10-067-514-1	Sequence 1, Appl:
	24	186.4	18.6	13467	10	US-09-764-868-1330	Sequence 1330, Ap
	25	186.4	18.6	13467	10	US-09-764-868-1498	Sequence 1498, Ap
	26	186.4	18.6	13467	11	US-09-764-891-8632	Sequence 8632, Ap
	27	185.8	18.6	790	13	US-10-027-632-145713	Sequence 145713,
	28	185.8	18.6	32190	9	US-09-764-869-2209	Sequence 2209, Ap
	29	185.8	18.6	32190	14	US-10-091-504-2309	Sequence 2309, Ap
	30	185.6	18.5	13058	10	US-09-764-846-303	Sequence 303, App
	31	185.6	18.5	13058	14	US-10-091-483-353	Sequence 303, App
	32	185.6	18.5	25001	12	US-10-017-161-2363	Sequence 2063, Ap
	33	185.6	18.5	62804	13	US-10-096-960-3	Sequence 3, Appl:
	34	185.2	18.5	32246	9	US-09-764-860-902	Sequence 802, App
	35	185.2	18.5	32246	14	US-10-074-095-952	Sequence 802, App
	36	185.2	18.5	41936	10	US-09-967-768A-116	Sequence 116, App
	37	185.2	18.5	162025	12	US-10-272-665-35	Sequence 35, App:
	38	185.2	18.5	162025	12	US-10-272-665-36	Sequence 36, App:
	39	185.2	18.5	162025	12	US-10-273-321-35	Sequence 35, Appl
	40	185.2	18.5	162025	12	US-10-273-321-36	Sequence 36, Appl
	41	185.2	18.5	162025	12	US-10-272-756-35	Sequence 35, Appl
	42	185.2	18.5	162025	12	US-10-272-756-36	Sequence 36, Appl
	43	185	18.5	421	11	US-09-818-995-4270	Sequence 4270, Ap
	44	185	18.5	637	13	US-10-027-632-218180	Sequence 218180,
	45	185	18.5	108317	12	US-10-017-161-2143	Sequence 2143, Ap

ALIGNMENTS

RESULT 1
US-10-275-910-1
; Sequence 1, Application US/10275910
; Publication No. US20030166142A1
; GENERAL INFORMATION:
; APPLICANT: Ramakrishnar, Shyam
; TITLE OF INVENTION: REGULATION OF HUMAN P2Y-LIKE G PROTEIN-COUPLED RECEPTOR
; FILE REFERENCE: 4974.03885
; CURRENT APPLICATION NUMBER: US/10/275,910
; CURRENT FILING DATE: 2002-11-12
; PRIOR APPLICATION NUMBER: US 60/203,582
; PRIOR FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: US 60/269,857
; PRIOR FILING DATE: 2001-02-21
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 3143
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: gene
; LOCATIONS: (520)...(1373)
US-10-275-910-1

Query Match	19.1%	Score 191.4	DB 12	Length 3143
Best Local Similarity	88.8%	Pred No. 21e-28		
Matches 207	Conservative 0	Mismatches 26	Indels 0	Gaps 0
QY	13	CATACCTTAGGCTGGGCAAGTGGCTCAGCTCTGTAAATCCAGCACTTTGGAGGCCAAG	72	
DB	2723	CATTATCTAGGCTGGGCAAGTGGCTCAGCTCTGTAAATCCAGCACTTTGGAGGCCAAG	2762	
QY	72	GTGGCGGATCACCTCAGGTTAGAGTTTACAGCAACCTGGGCATCATGGCGAAACCT	132	
DB	2783	GCGGGCGGATCACCTCAGGTTAGAGTTTACAGCAACCTGGGCATCATGGCGAAACCT	2842	
QY	133	GTCTCTACTAAATAACAAAATTAGCTGGGCATGGAGGCATGTGCTGTATCCAGCT	192	

```
Db      2843 ATCTCTACTAAAAATACAAAAAATTAGCTGGGCGTGGTGGCGGTGGCTATATATCCAGCT 2902
QY      193 ACTGGGAGGCTGAGACAGAGAATTGTTGAACCTTTGGAGGGGGAGGTTGCA 245
EB      2933 ACTGGGAGGCTGAGAAAGGAGATCGCTTGAACCCGGGAGGGCGGAGGTTGCA 2955

RESULT 2
US-09-764-891-6869
; Sequence 6869, Application US/09764891
; Publication No. US20030377808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-02-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6869
; LENGTH: 30573
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-6869

Query Match      19.1%; Score 191.4; DB 11; Length 30573;
Best Local Similarity 90.7%; Pred. No. 5.1e-28;
Matches 204; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY      21 AGGCTGGGCAAGTGGCTCAGCTGTGTAATCCAGCACTTTGGAGGGCAAGGTGGGGG 80
DB      6984 AGGCTGGGCAAGTGGCTCAGCTGTGTAATCCAGCACTTTGGAGGGTGGAGGTAGGTGG 7043

QY      8: ATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAAACCCCTGTCTTAC 140
DB      7344 ATCATCTGAGTTCAGGAGTTCAGGACCAACCTGGCCATCATGGTGAACCCCTGTCTTAC 7103

QY      141 TAAAAATACAAAAATTAGCTGGGCAATGGAGGCATGTGCTGTATATCCAGCTACTCGGGA 200
DB      7104 TAAAAATACACAATTAGCTGGGCAATGGTGGCAGCTGCTGTATATCCAGCTACTTGGGA 7163

QY      201 GGCTGAGACAGAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
DB      7164 GGCTGAGGCAAGGAGATTGCTTGAACCTTGGAGGGTGGAGGTTGCA 7208

RESULT 3
US-09-764-891-6868
; Sequence 6868, Application US/09764891
; Publication No. US20030377808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6868
; LENGTH: 32195
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-6868

Query Match      19.1%; Score 191.4; DB 11; Length 32195;
Best Local Similarity 90.7%; Pred. No. 5.3e-28;
Matches 204; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY      21 AGGCTGGGCAAGTGGCTCAGCTGTGTAATCCAGCACTTTGGAGGGCGCAAGGTGGGGG 80
DB      12766 AGGCTGGGCAAGTGGCTCAGCTGTGTAATCCAGCACTTTGGAGGGTGGAGTTGG 12825
```

```
QY      81 ATCACCTGAGGTTAGGAGTTCAGGACCAACCTGGCCATCATGGCGAAACCCCTGTCTTAC 140
DB      12826 ATCATCTGAGTTCAGGAGTTCAGGACCAACCTGGCCATCATGGTGAACCCCTGTCTTAC 12825

QY      141 TAAAAATACAAAAATTAGCTGGGCAATGGAGGCATGTGCTGTATATCCAGCTACTCGGGA 200
DB      12886 TAAAAATACACAATTAGCTGGGCAATGGTGGCAGCTGCTGTATATCCAGCTACTTGGGA 12945

QY      201 GGCTGAGACAGAGAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
DB      12946 GGCTGAGGCAAGGAGATTGCTTGAACCTTGGAGGTGGAGGTTGCA 12990

RESULT 4
US-09-814-353-21993
; Sequence 21993, Application US/09814353
; Publication No. US20030165831A1
; GENERAL INFORMATION:
; APPLICANT: Lee, John
; APPLICANT: Thompson, Pamela
; APPLICANT: Lillie, James
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF OVARIAN CANCER
; FILE REFERENCE: MRI-006B
; CURRENT APPLICATION NUMBER: US/09/814,353
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/191,031
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: US 60/207,124
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 60/211,940
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: US 60/216,820
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: US 60/220,661
; PRIOR FILING DATE: 2000-07-25
; PRIOR APPLICATION NUMBER: US 60/257,672
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 22037
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21993
; LENGTH: 7677
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1, 2, 3, 7675, 7676, 7677
; OTHER INFORMATION: n = A,T,C or G
US-09-814-353-21993

Query Match      19.1%; Score 190.8; DB 12; Length 7677;
Best Local Similarity 60.7%; Pred. No. 3.9e-28;
Matches 368; Conservative 0; Mismatches 227; Indels 11; Gaps 3;

QY      1 AATATAAAATATGATATCTTAGGCTGGGCAAGTGGCTCACGCTGTGTATATCCAGCACTT 60
DB      6836 AATTATTAGAAATTAACTCGGCTGGGCGGGTGGCTCACGCTGTGTATATCCAGCACTT 6895

QY      61 TGGAGGGCCCAAGGTGGGCGGATCACCTGAGGTTAGGAGTTAGGACCAACCTGGCCCATCA 120
DB      6896 TGGGAAGCCCAAGGTGGGCGGATCACTTCAGGTTAGGAGTGGAGACCAAGCTGGCCCAACA 6955

QY      121 TGGCGAAACCCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCAATGGAGGCATGTGCCT 180
DB      6956 TGGCGAAACCCCTGTCTCTACTAAAAATACAAAAATTAGCCGGGCGTGGTGGCACTTGCCT 7015

QY      181 GTAATCCAGCTACTCGGGAGGCTGAGACAGAGAATTGCTTGAACCTTGGAGGGGGAGG 240
DB      7016 GTAATCCAGCTACTCGGGAGGCTGAGGACAGAGAAATCGATTAAACCTTGGAGGGCGGAGA 7075

QY      241 TTGC----ATATCTGAGTGGTGAATTTGTATCTTTTCTCTTTTCTCTTTTCTGTATTTTG 296
```


Db 7076 TTGCAATGAGATCGAACCCACTGCACTCCAGCTGGGTGACAGAGAGAGCACTGCCCTGG 7135
QY 297 AACITTTCTATAAATGATTGTGTTTGTATATTGGGAAATATTAGCTTTCAAA 356
Db 7136 AAAAAAAGAAATCTCACTCACTATCTAGAGAGGATTGTGAGAAATATTACGATTCCGGT 7195
QY 357 GTTAATACCTATGAACCTAAACACAAAGTAATTAATATATAGTATAGCATTTATTAGGT 416
Db 7196 CTTGAACCTTGAATTATGCAAAAGAGGTATATAATTAATTTTCATTATGATTCAGTTT 7255
QY 417 TTCTTGTTAGAGATCAACATAGAAATATATTTAAATGGCTGACATATTCTTAAAG 476
Db 7256 TTAAGGCTTTGCAGCT--TCTATAGTGTCTCAGATGGCACTAGATATTTTAAAGC 7312
QY 477 ATACATACAGTATATTTTATTAACATTA---AGAAACAGCAGCATGATTACTTTAAT 532
Db 7313 ATCATATTAGAAATACTTTTAAAGAGACTTATATAGAAATAGAGATTCTGATTTTAC 7372
QY 533 CATCATTTGCTTAACCATATATACCTGTGATCATTTGTATTTGTCATGCTGCTTTTAA 592
Db 7373 AGNGGATTTGCTTCAATTAAGCCAGATTCTGTAAAGTTTTCATTTGAAATTTCTAGTAA 7432
QY 593 AATCT 598
Db 7433 ACATAT 7439

RESULT 5
US-09-836-613-3
Sequence 3, Application US/09836613
Publication No. US20030039643A
GENERAL INFORMATION:
APPLICANT: HOPWOOD, JOHN JOSEPH; SCOTT, HAMISH STEELE;
WEBER, BIRGIT; BLANCH, LIAMNE; ANSON, DONALD STEWART
TITLE OF INVENTION: SYNTHETIC MAXYALIAN
N-ACETYLGLUCOSAMINIDASE AND GENETIC SEQUENCES
ENCODING SAME
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON PEABODY LLP
STREET: 99C STEWART AVENUE
CITY: GARDEN CITY
STATE: NEW YORK
COUNTRY: UNITED STATES
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/836,613
FILING DATE: 17-Apr-2001
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/00747
FILING DATE: 22-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: POKALSKY, ANN S.
REGISTRATION NUMBER: 34,697
REFERENCE/DOCKET NUMBER: 22497/104
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516 742 4343
TELEFAX: 516 742 4366
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 10380 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Homo sapiens

POSITION IN GENOME:
CHROMOSOME/SEGMENT: Chromosome 17
FEATURE:
NAME/KEY: exon 1
LOCATION: 990..1372
FEATURE:
NAME/KEY: exon 2
LOCATION: 2115..2262
FEATURE:
NAME/KEY: exon 3
LOCATION: 3056..3202
FEATURE:
NAME/KEY: exon 4
LOCATION: 3387..3472
FEATURE:
NAME/KEY: exon 5
LOCATION: 5667..5923
FEATURE:
NAME/KEY: exon 6
LOCATION: 7745..8955
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-836-613-3
Query Match 19.0%; Score 189.8; DB 1; Length 10380;
Best Local Similarity 86.7%; Pred. No. 7e-28;
Matches 209; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
QY 5 TAAAAATGCATACTTTTAGGCTGGGCAAAAGTGGCTCAGCTGTGTATATCCACGCACTTTGGG 64
Db 3540 TAAAAACTTAAGCTCTGGGCGGGCGGAGTGGCTCAGCTGTGTATATCCACGCACTTTGGG 3599
QY 65 AGGCCAAGGTGGGCGGATCAGCTAGGTTAGGAGTTCAGGACCAACCTGGGCAATCATGGC 124
Db 3600 AGGCCAGATTGGGCGGATCAGCTAGGTTAGGAGTTCAGGACCAACCTGGGCAATCATGGC 3659
QY 125 GAAACCTGTCTTACTTAAATAACAAATTAAGCTGGGATGGAGGCAATGTGCTGTAA 184
Db 3660 GAAACCTGTCTTACTTAAATAATATAAAATTAAGCTGGGATGGGCAATGTGCTGTAA 3719
QY 185 TCCAGCTACTGGGAGGCTGAGACAGAGAAATTCCTTGAACCTTGGAGGGGAGGTTGC 244
Db 3720 TCCATCTACTAGGAGGCTGAGACAGGAGAAATTCCTTAAACCTGGGAGGAGCGTTGC 3779
QY 245 A 245
Db 3780 A 3780
RESULT 6
US-10-027-632-123771
Sequence 123771, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 123771
; LENGTH: 1325
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-123771

Query Match 18.9%; Score 189.4; DB 13; Length 1325;
Best Local Similarity 86.3%; Pred. No. 3.8e-28;
Matches 208; Conservative 1; Mismatches 32; Indels 0; Gaps 0;

QY 4 ATAAAAATGCATCTTTAGGCTGGGCAAGTGGCTCAGCTCTGTAAATCCAGCAGCTTTGG 63
|||
DB 206 ATTAAAAAAAWTTTTTGGCTGGGCACGGTGCTCACACCTGTAAATCCAGCAGCTTTGG 265

QY 64 GAGGCCAAGGTGGCGGATCAGCTGAGGTTAGGAGTTCAGGAGTTCAGGACCAACCTGGCCATCATGG 123
|||
DB 266 GAGGCCAAGGTGGGTGGATTACTGAGGTCAGGAGTTCGAGACCAAGCCTGGCCAACTGG 325

QY 124 CGAAACCCCTGTCTCTACTAAATAACAAAAATTAGCTGGGATGGAGGTCATGTGCCTGTA 183
|||
DB 326 TGAACCCCTGCATCTACTTAAATAACAAAAATTAGCTGGGATGGAGGTCATGTGCCTGTA 385

QY 184 ATCCAGCTACTCGGAGGCTGAGACAGAGAAATTGCTTGAACCTTGGAGGGGAGGTTG 243
|||
DB 386 ATCCAGCTACTCAGGAGGCTGAGGAGGAGGAGGAAATCATTGAACCTGGAGGGGAGGTTG 445

QY 244 C 244
|
DB 446 C 446

RESULT 7

US-10-027-632-123772
; Sequence 123772, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 10827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,005
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 123772
; LENGTH: 1325
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-123772

Query Match 18.9%; Score 189.4; DB 13; Length 1325;
Best Local Similarity 86.3%; Pred. No. 3.8e-28;
Matches 208; Conservative 1; Mismatches 32; Indels 0; Gaps 0;

QY 4 ATAAAAATGCATCTTTAGGCTGGGCAAGTGGCTCAGCTCTGTAAATCCAGCAGCTTTGG 63
|||
DB 206 ATTAAAAAAAWTTTTTGGCTGGGCACGGTGCTCACACCTGTAAATCCAGCAGCTTTGG 265

QY 64 GAGGCCAAGGTGGCGGATCAGCTGAGGTTAGGAGTTCAGGAGTTCAGGACCAACCTGGCCATCATGG 123
|||

DB 266 GAGGCCAAGGTGGGTGGATTACTGAGGTCAGGAGTTCGAGACCAACCTGGCCAACTGG 325

QY 124 CGAAACCCCTGTCTCTACTAAATAACAAAAATTAGCTGGGATGGAGGTCATGTGCCTGTA 183
|||
DB 326 TGAACCCCTGCATCTACTTAAATAACAAAAATTAGCTGGGATGGAGGTCATGTGCCTGTA 385

QY 184 ATCCAGCTACTCGGAGGCTGAGACAGAGAAATTGCTTGAACCTTGGAGGGGAGGTTG 243
|||
DB 386 ATCCAGCTACTCAGGAGGCTGAGGAGGAGGAGGAAATCATTGAACCTGGAGGGGAGGTTG 445

QY 244 C 244
|
DB 446 C 446

RESULT 8

US-10-027-632-123773
; Sequence 123773, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 10827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 123773
; LENGTH: 1325
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-123773

Query Match 18.9%; Score 189.4; DB 13; Length 1325;
Best Local Similarity 86.3%; Pred. No. 3.8e-28;
Matches 208; Conservative 1; Mismatches 32; Indels 0; Gaps 0;

QY 4 ATAAAAATGCATCTTTAGGCTGGGCAAGTGGCTCAGCTCTGTAAATCCAGCAGCTTTGG 63
|||
DB 206 ATTAAAAAAAWTTTTTGGCTGGGCACGGTGCTCACACCTGTAAATCCAGCAGCTTTGG 265

QY 64 GAGGCCAAGGTGGCGGATCAGCTGAGGTTAGGAGTTCAGGAGTTCAGGACCAACCTGGCCATCATGG 123
|||
DB 266 GAGGCCAAGGTGGGTGGATTACTGAGGTCAGGAGTTCGAGACCAACCTGGCCAACTGG 325

QY 124 CGAAACCCCTGTCTCTACTAAATAACAAAAATTAGCTGGGATGGAGGTCATGTGCCTGTA 183
|||
DB 326 TGAACCCCTGCATCTACTTAAATAACAAAAATTAGCTGGGATGGAGGTCATGTGCCTGTA 385

QY 184 ATCCAGCTACTCGGAGGCTGAGACAGAGAAATTGCTTGAACCTTGGAGGGGAGGTTG 243
|||
DB 386 ATCCAGCTACTCAGGAGGCTGAGGAGGAGGAGGAAATCATTGAACCTGGAGGGGAGGTTG 445

QY 244 C 244
|
DB 446 C 446

```
RESULT 9
US-10-027-632-102144/c
; Sequence 102144, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 10827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 102144
; LENGTH: 2490
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-102144
Query Match      18.9%   Score 189.4;   DB 13;   Length 2490;
Best Local Similarity 87.0%;   Pred. No. 4.8e-28;
Matches 209;   Conservative 0;   Mismatches 31;   Indels 0;   Gaps 0;

QY 6 AAAAAATGCATCTTTAGGCTGGGCAAGTGGCTCAGCTGTGTATCCAGCACCTTTGGGA 65
DB 476 AAAAAAATAAAGGTGGGCAATGGTCTCAGCTGTGTATCCAGCACCTTTGGGA 417
QY 66 GGCCAGGTGGGCGGATCAGCTGAGGTAGGATTCAGGACCAACCTGGCCATCATGGC 125
DB 416 GGCCAGGTGGGCGGATCAGCTGAGGTAGGATTCAGGACCAACCTGGCCATCATGGC 357
QY 126 AAACCCCTGTCTTACTAAAAATACAAAAATAGCTGGGATGGAGGATGTGCTGTAAAT 185
DB 356 AAACCCCTGTCTTACTAAAAATACAAAAATAGCTGGGATGGAGGATGTGCTGTAAAT 297
QY 186 CCCAGCTACTCGGAGGCTGAGGACAGAGAAATTCCTTGAACCTTGGAGGGGAGGTGC 244
DB 296 CCCAGCTACTCGGAGGCTGAGGACAGAGAAATTCCTTGAACCTTGGAGGGGAGGTGC 238

RESULT 10
US-10-027-632-111607
; Sequence 111607, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 10827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
```

```

; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 111607
; LENGTH: 2490
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-111607
Query Match      18.9%   Score 189.4;   DB 13;   Length 2490;
Best Local Similarity 87.0%;   Pred. No. 4.8e-28;
Matches 208;   Conservative 0;   Mismatches 31;   Indels 0;   Gaps 0;

QY 6 AAAAAATGCATCTTTAGGCTGGGCAAGTGGCTCAGCTGTGTATCCAGCACCTTTGGGA 65
DB 2015 AAAAAAATAAAGGTGGGCAATGGTCTCAGCTGTGTATCCAGCACCTTTGGGA 2074
QY 66 GGCCAGGTGGGCGGATCAGCTGAGGTAGGATTCAGGACCAACCTGGCCATCATGGC 125
DB 2075 GGCCAGGTGGGCGGATCAGCTGAGGTAGGATTCAGGACCAACCTGGCCATCATGGC 2134
QY 126 AAACCCCTGTCTTACTAAAAATACAAAAATAGCTGGGATGGAGGATGTGCTGTAAAT 185
DB 2135 AAACCCCTGTCTTACTAAAAATACAAAAATAGCTGGGATGGAGGATGTGCTGTAAAT 2194
QY 186 CCCAGCTACTCGGAGGCTGAGGACAGAGAAATTCCTTGAACCTTGGAGGGGAGGTGC 244
DB 2195 CCCAGCTACTCGGAGGCTGAGGACAGAGAAATTCCTTGAACCTTGGAGGGGAGGTGC 2253

RESULT 11
US-10-027-632-306904
; Sequence 306904, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
; FILE REFERENCE: 10827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 306904
; LENGTH: 614
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-306904
Query Match      18.9%   Score 188.8;   DB 13;   Length 614;
Best Local Similarity 86.4%;   Pred. No. 3.7e-28;
Matches 205;   Conservative 0;   Mismatches 27;   Indels 0;   Gaps 0;

QY 14 ATACTTTAGGCTGGGCAAGTGGCTCAGCTGTGTATCCAGCACCTTTGGGAGGCAAGG 73
DB 244 ATATTAGGCTGGGCAAGTGGCTCAGCTGTGTATCCAGCACCTTTGGGAGGCAAGG 303
```



```
QY      6  AAAAAATGCATACCTTTAGGCTGGGCAAAAGTGGCTCACGTCTGTATCCAGCACCTTTGGGA 65
      |||||  ||  |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1340: AAAAAATCACTCTTTCCGGTGGGGCGAGTGGCTCACACCTGTATCCAGCACCTTTGGGA 13460
      |||||  ||  |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      66  GGCCAAAGGTGGCGGATCACCTGAGGTAGGAGTTAGGACCAACCTGGCCATCATGGCG 125
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 13461: GGCCAAAGGTGGCGGATCACCTGAGGTAGGAGTTAGGACCAACCTGGCCATCATGGTG 13520
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      126  AAACCCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAA 185
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 13521: AAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGTGTGTGGACACACACCTGTAA 13580
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      186  CCCAGCTACTGGGAGGTGAGACAGAGAAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 13581: CCCAGCTACTGGGAGGTGAGGCAAGAGAAATTGCTTGAACCTTGGAGATGGAGGTTGCA 13640
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

RESULT 15

```
US-10-091-504-1599
; Sequence 1599, Application US/10091504
; Publication No. US2003005908A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al
;   TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
;   FILE REFERENCE: PC007C1
;   CURRENT APPLICATION NUMBER: US/10/091,504
;   CURRENT FILING DATE: 2002-03-07
;   NUMBER OF SEQ ID NOS: 2442
;   Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1599
;   LENGTH: 17792
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-091-504-1599
```

```
Query Match      18.9%; Score 188.8; DB 14; Length 17792;
Best Local Similarity 66.7%; Pred. No. 1.4e-27;
Matches 209; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
```

```
QY      6  AAAAAATGCATACCTTTAGGCTGGGCAAAAGTGGCTCACGTCTGTATCCAGCACCTTTGGGA 65
      |||||  ||  |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1340: AAAAAATCACTCTTTCCGGTGGGGCGAGTGGCTCACACCTGTATCCAGCACCTTTGGGA 13460
      |||||  ||  |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      66  GGCCAAAGGTGGCGGATCACCTGAGGTAGGAGTTAGGACCAACCTGGCCATCATGGCG 125
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 13461: GGCCAAAGGTGGCGGATCACCTGAGGTAGGAGTTAGGACCAACCTGGCCATCATGGTG 13520
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      126  AAACCCCTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGCATGTGCCTGTAA 185
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 13521: AAACCCCATCTCTACTAAAAATACAAAAATTAGCTGGGTGTGTGGACACACACCTGTAA 13580
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY      186  CCCAGCTACTGGGAGGTGAGACAGAGAAATTGCTTGAACCTTGGAGGGGGAGGTTGCA 245
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 13581: CCCAGCTACTGGGAGGTGAGGCAAGAGAAATTGCTTGAACCTTGGAGATGGAGGTTGCA 13640
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

Search completed: October 24, 2003, 18:34:54
Cpu time : 250.585 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 18:12:56 ; Search time 2091.3 Seconds
(without alignments)
11633.309 Million cell updates/sec

Title: US-09-830-902-1_COPY_100000_101000
Perfect score: 1001
Sequence: 1 aatataaaaatgcatacttt.....acagatctcgggtctcttt 1001

Scoring table: IDENTITY_NJC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 1215223056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: em_estha:
2: em_esthum:
3: em_estin:
4: em_estmu:
5: em_estov:
6: em_estpl:
7: em_estro:
8: em_estci:
9: gb_est1:
10: gb_est2:
11: gb_est3:
12: gb_est4:
13: gb_est5:
14: gb_est6:
15: em_estfun:
16: em_estom:
17: em_gss_hum:
18: em_gss_inv:
19: em_gss_pln:
20: em_gss_vrt:
21: em_gss_fun:
22: em_gss_mam:
23: em_gss_mus:
24: em_gss_pro:
25: em_gss_rtd:
26: em_gss_phg:
27: em_gss_vrl:
28: gb_gss1:
29: gb_gss2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	530.4	53.0	556	14	CB241186
2	511.2	51.1	532	9	AW297897
3	403	40.3	743	13	BJ853405
4	398.2	39.8	877	13	BJ193776

5	388.6	38.8	784	10	BG204678
6	388	38.8	749	10	BG217901
7	350	35.0	694	10	BG193969
8	286.2	28.6	728	10	BF216343
9	277.8	27.8	260	10	BG195583
10	244.6	24.4	460	10	BE172740
11	215.2	21.5	239	10	BG208384
12	209.4	20.9	401	14	W42916
13	190.4	19.0	516	28	AQ112451
14	189.8	18.9	515	14	CA946732
15	187.4	18.7	446	9	A1339725
16	187.4	18.7	469	10	BG574581
17	187.4	18.7	660	9	A1708207
18	187.2	18.7	660	28	AQ057239
19	187.2	18.7	711	28	AQ415030
20	187	18.7	499	2	HSM036977
21	186.8	18.7	363	9	AA493808
22	186.6	18.6	500	9	AW338860
23	186.6	18.6	826	13	BQ438665
24	186.4	18.6	400	2	HSM035126
25	186.4	18.6	410	28	AQ673107
26	186.4	18.6	794	12	BM679479
27	186.2	18.6	368	9	AA603264
28	186.2	18.6	397	9	A1821805
29	186.2	18.6	407	9	A1821056
30	186.2	18.6	425	9	A1792092
31	186.2	18.6	439	9	A1791718
32	186.2	18.6	442	28	AQ018157
33	186.2	18.6	600	13	BX093702
34	186.2	18.6	707	28	AQ540344
35	185.8	18.6	544	9	AL699310
36	185.6	18.5	443	10	BF991881
37	185.6	18.5	583	28	AQ391420
38	185.6	18.5	662	29	AG037594
39	185.4	18.5	361	9	AW265688
40	185.4	18.5	700	28	AQ317190
41	185.4	18.5	735	29	AG094268
42	185.2	18.5	435	10	BF991882
43	185.2	18.5	607	29	BZ603705
44	185	18.5	365	9	AA278496
45	184.8	18.5	342	9	AA846923

ALIGNMENTS

RESULT :
CB241186
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
COMMENT

CB241186 556 bp mRNA linear EST 12-FEB-2003
UI-CF-FNC-afx-b-20-C-UI-51 UI-CF-FNO Homo sapiens cDNA clone
UI-CF-FNO-afx-b-20-C-UI 3', mRNA sequence.
CB241186
CB241186.1 GI:28362830
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. bases 1 to 556
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9): 791-806 (1996)
97044477
5889548
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
E-mail: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa


```
BASE COUNT      194 a      59 c      77 g      172 t
ORIGIN

Query Match      51.1%; Score 511.2; DB 9; Length 532;
Best Local Similarity 99.4%; Pred. No. 1.9e-06;
Matches 513; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 391 TATATTAGTATAGCACTTTATTAAAGGTTTCTTGTGTAGCAGATCAACATAGAAATATATT 450
DB 17 TAAATAAGTATAGCACTTTATTAAAGGTTTCTTGTGTAGCAGATCAACATAGAAATATATT 76
QY 451 TAAATGGCTGACATAAATTTCTAAGATATACATACACGGTATATTTTATATACATTAGAA 510
DB 77 TAAATGGCTGACATAAATTTCTAAGATATACATACACGGTATATTTTATATACATTAGAA 136
QY 511 ACAGCAGCATCACTTAACTTAATCCATCATCTTGGTTAACGACCATATACCGTTGATCACT 570
DB 137 ACAGCAGCATCACTTAACTTAATCCATCATCTTGGTTAACGACCATATACCGTTGATCACT 196
QY 571 TGTATTGTGATGTGCTTTTAAATATCTAGATGAGAAATATTCGATTATCTGACTTCACT 630
DB 197 TGTATTGTGATGTGCTTTTAAATATCTAGATGAGAAATATTCGATTATCTGACTTCACT 256
QY 631 GAATCCTTGAAAAAATAAAGCGGAGCGGTGACCCCTCAAACTTTAGAGCGGTACATAGCT 690
DB 257 GAATCCTTGAAAAAATAAAGCGGAGCGGTGACCCCTCAAACTTTAGAGCGGTACATAGCT 316
QY 691 TGGACACAGGACTTTGGAGATACCACTGTTTAAAGGAATATACCTTGTAAACCTGAGAAC 750
DB 317 TGGACACAGGACTTTGGAGATACCACTGTTTAAAGGAATATACCTTGTAAACCTGAGAAC 376
QY 751 ATTTTACTTAAAGAGGAAACACAGATCTTCAATGAGGATCTGCGGTACAGAAACAGC 810
DB 377 ATTTTACTTAAAGAGGAAACACAGATCTTCAATGAGGATCTGCGGTACAGAAACAGC 436
QY 811 CTAAGTTTACAGGACTTTTGTAGAGTCTTACATATTTTGTGACCCAACTTGAAGATGACC 870
DB 437 CTAAGTTTACAGGACTTTTGTAGAGTCTTACATATTTTGTGACCCAACTTGAAGATGACC 496
QY 871 AGAAACAGACTTAAACAAATATATACAAATGCAATG 906
DB 497 AGAAACAGACTTAAACAAATATATACAAATGCAATG 532

RESULT 3
BU853405
LOCUS
DEFINITION
AGENCOURT_10418018 NIH_MGC_82 Homo sapiens cDNA clone IMAGE:6620245
5', mRNA sequence.
BU853405
VERSION
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 743)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: CLONTECH
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LJCX2870 row: a column: 13
High quality sequence stop: 646.
Location/Qualifiers
1..743
source
```

```
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6620245"
/lab_host="DH10B (T3 phage-resistant)"
/clone_lib="NIH_MGC_82"
/notes="Organ: testis; Vector: pDNR-LIB (Clontech); Site 1:
SfiI (ggcgctcgcc); Site 2: SfiI (ggcattatggcc); 5' and
3' adaptors were used in cloning as follows: 5' adaptor
sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
5'-ATTCTAGAGCGGAGCGGCGGACATG-dT(30)BN-3' where B = A,
C, or G and N = A, C, G, or T; Average insert size
1.35 kb (range 0.9-4.0 kb); 14/15 colonies contained
inserts by PCR. This library was enriched for full-length
clones and was constructed by Clontech Laboratories (Palo
Alto, CA)."
```

BASE COUNT 248 a 123 c 121 g 250 t : others

Query Match 40.3%; Score 403; DB 13; Length 743;
Best Local Similarity 100.0%; Pred. No. 0.0012;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 599 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAATAAAGCGGCG 658
DB 11 AGATGAGAAATATTCGATTATCTGACTTCACTGAATCCTTGAAAAAATAAAGCGGCG 70

QY 659 TCAGCCCTCAAACTTTAGAGCGGTACATACCTTGGACAGGAGACTTTGGAGATACCACTG 718
DB 71 TCAGCCCTCAAACTTTAGAGCGGTACATACCTTGGACAGGAGACTTTGGAGATACCACTG 130

QY 719 TTTAAGGAATATACCTTTGTAAACCTGCGAGACATTTTACTTAAAGAGGAAACACAGAT 778
DB 121 TTTAAGGAATATACCTTTGTAAACCTGCGAGACATTTTACTTAAAGAGGAAACACAGAT 190

QY 779 CTTCAATGAGCGTCATCGGCTACAGAAACAGCCCTAGTTTACAGGACTTTTAGAGTCTT 838
DB 191 CTTCAATGAGCGTCATCGGCTACAGAAACAGCCCTAGTTTACAGGACTTTTAGAGTCTT 250

QY 839 ACATATTTGTCACCAAACTTGAAGATGAACCCAGAAAAACAGACTTAAACAAATATACAA 898
DB 251 ACATATTTGTCACCAAACTTGAAGATGAACCCAGAAAAACAGACTTAAACAAATATACAA 310

QY 899 TCGAAATGTAAATTTTGTGTTTAAAGGCTTGCCTTGATGGTGCACAGTTATCCCAATGG 958
DB 311 TCGAAATGTAAATTTTGTGTTTAAAGGCTTGCCTTGATGGTGCACAGTTATCCCAATGG 370

QY 959 ACACTAAGTTAGAGCACACAAACCTGATTTCTGGTCTTCTTT 1001
DB 371 ACACTAAGTTAGAGCACACAAACCTGATTTCTGGTCTTCTTT 413

RESULT 4
BU193776
LOCUS
DEFINITION
AGENCOURT_7905132 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6161142
5', mRNA sequence.
BU193776
VERSION
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 877)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
http://image.llnl.gov
Plate: LLAM13512 row: p column: 07
High quality sequence stop: 579.
Location/Qualifiers

FEATURES

source
1..877
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6161142"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="N3H MGC 72"
/note="Organ: skin; Vector: pCMV-SPORT6; Site 1: Not;
Site 2: Salt; Cloned unidirectionally. Primer: CligedT.
Average insert size 2 kb. Library constructed by Life
Technologies."
300 a 166 c 177 g 234 t

Query Match: 33.8%; Score 396.2; DB 13; Length 877;
Best Local Similarity 99.3%; Pred. No. 0.0015;
Matches 400; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 599 AGATGAGAAATATTCGATTATCTGACTTCACCTGAAATCCTTGAAAAAATAAAACGCAGCG 658
D5 293 AGATGAGAAATATTCGATTATCTGACTTCACCTGAAATCCTTGAAAAAATAAAACGCAGCG 352
QY 659 TCAGCCCTCAAACTTTAGAACGCTACATACCTTGGAAACAGGACTTTGGAGTACCACCTG 718
D5 353 TCAGCCCTCAAACTTTAGAACGCTACATACCTTGGAAACAGGACTTTGGAGTACCACCTG 412
QY 719 TTTRAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAGAGGAAACACAGAT 778
D5 413 TTTRAGGAAATACCTTTGTAAACCTGCAGAACATTTTACTTAAAGAGGAAACACAGAT 472
QY 779 CTGCAATGACGTCATCGGCTACAGAACAGCCTAAGTTTACAGGACTTTTACAGTCTT 838
D5 473 CTGCAATGACGTCATCGGCTACAGAACAGCCTAAGTTTACAGGACTTTTACAGTCTT 532
QY 839 ACATATTGTGCACCAAACTTGAAGATGAACCAAGAACAGACTTAAACAAATATACAA 898
D5 533 ACATATTGTGCACCAAACTTGAAGATGAACCAAGAACAGACTTAAACAAATATACAA 592
QY 899 TGCAATGTAAATTTTGTGTTTAAAGCCCTTGCCTTGCATGCTGACACATTTACCAATGG 958
D5 593 TGCAATGTAAATTTTGTGTTTAAAGCCCTTGCCTTGCATGCTGACACATTTACCAATGG 652
QY 959 ACACCTAAGTTAGGACACACAAACCTGATTCCTGCTCTTT 1001
D5 653 ACACCTAAGTTAGGACACACAAACCTGATTCCTGCTCTTT 695

RESULT 5
BG204678
LOCUS
DEFINITION
RST24093 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
EST 21-APR-2001
ACCESSION
BG204678
VERSION
BG204678.1 GI:13726365
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 784)
AUTHORS
Harrington, J.C., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whittington, J.,
Lerner, L., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith,
E., Veloso, N., Klika, A., Hess, J., Cochren, K., Lo, K., Offenbacher,
J., Danzig, J. and Ducar, M.
Creation of genome-wide protein expression libraries using random

activation of gene expression
Nat. Biotechnol. 19 (5), 440-445 (2001)
2122715;
PUBMED
11329013
COMMENT
Contact: Scott C. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@atersys.com
High quality sequence stop: 500.
Location/Qualifiers

FEATURES

source
1..784
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_line="HT1080"
/clone_lib="Athersys RAGE Library"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
255 a 141 c 129 g 259 t

BASE COUNT 255 a 141 c 129 g 259 t
ORIGIN
Query Match: 38.8%; Score 388.6; DB 10; Length 784;
Best Local Similarity 99.0%; Pred. No. 0.0029;
Matches 391; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 607 AATATTCGATTATCTGACTTCACCTGAAATCCTTGAAAAAATAAAACGCAGCGCTCAGCCCT 666
D5 7 AATATTCGATTATCTGACTTCACCTGAAATCCTTGAAAAAATAAAACGCAGCGCTCAGCCCT 66
QY 667 CAACTTTAGAACGCTACATACCTTGGAAACAGGACTTTGGAGATACCACCTGTTTAAAGGA 726
D5 67 CATACTTTAGAACGCTACATACCTTGGAAACAGGACTTTGGAGATACCACCTGTTTAAAGGA 126
QY 727 AATACCTTTGTAAACCTGCAGAACATTTTACTTAAAGAGGAAACACAGATCTTCAATG 786
D5 127 AATACCTTTGTAAACCTGCAGAACATTTTACTTAAAGAGGAAACACAGATCTTCAATG 186
QY 787 AACGTCATCGGCTACGAAACAGCCCTAAGTTTACAGGACTTTTACAGTCTTACATATTT 846
D5 187 AACGTCATCGGCTACGAAACAGCCCTAAGTTTACAGGACTTTTACAGTCTTACATATTT 246
QY 847 GTGCACCAAACTTGAAGATGAACCAAGAACAGACTTAAACAAATATACAAATGCAATG 906
D5 247 GTGCACCAAACTTGAAGATGAACCAAGAACAGACTTATACAAATATACAAATGCAATG 306
QY 907 TAAATTTTGTGTTTAAAGCCCTTGCCTTGCATGCTGACAGTTATCCCAATGGACACTAAG 966
D5 307 TAAATTTTGTGTTTAAAGCCCTTGCCTTGCATGCTGACAGTTATCCCAATGGACACTAAG 366
QY 967 TTAGAGCACACAAACCTGATTCCTGCTCTTT 1001
D5 367 TTAGAGCACACAAACCTGATTCCTGCTCTTT 401

RESULT 6
BG217901
LOCUS
DEFINITION
RST37625 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
EST 21-APR-2001
ACCESSION
BG217901
VERSION
BG217901.1 GI:13743922
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 749)
AUTHORS
Harrington, J.C., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,

Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whittington, J., Lerner, S., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith, E., Veloso, N., Klika, A., Hess, J., Cothren, K., So, K., Offenbacher, J., Danzig, J., and Ducar, M.

Creation of genome-wide protein expression libraries using random activation of gene expression

Nat. Biotechnol. 19 (5), 440-445 (2001)

21227151

11329013

Contact: Scott C. Cain

Athersys, Inc.

3201 Carnegie Ave, Cleveland, OH 44115, USA

Tel: 216 431 9900

Fax: 216 361 9596

Email: scaine@athersys.com

High quality sequence stop: 551

Location/Qualifiers

1..749

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/cell_line="HT1080"

/clone_lib="Athersys RAGE Library"

/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT 251 a 127 c 121 g 250 t

ORIGIN

Query Match 38.8%; Score 389; DB 10; Length 743;
Best Local Similarity 99.5%; Pred. No. 3.0031;
Matches 399; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 603 GAGAAATATTCGATTATC TGAATTCATCGAATCCTTGAAAAAATAAAGCGCGCTCA 661
Db 1 GAGAAATATTCGATTATC TGAATTCATCGAATCCTTGAAAAAATAAAGCGCGCTCA 60
QY 662 GCCCTCAACTTAGAGCGTACATACGTTGGACAGGACTTTGGAGATACCAGTCTT 721
Db 61 GCCCTCAACTTAGAGCGTACATACGTTGGACAGGACTTTGGAGATACCAGTCTT 120
QY 722 AAGGAAATACCTTTGTAAAGCTGCAGAACATTTTACTTAAAGAGGAAACACAGATCTT 781
Db 121 AAGGAAATACCTTTGTAAAGCTGCAGAACATTTTACTTAAAGAGGAAACACAGATCTT 180
QY 782 CAATGAACGTCATCGGCTACAGAAACAGCGCTTAAGTTTACAGGACTTTTGGAGTCTTACA 841
Db 181 CAATGAACGTCATCGGCTACAGAAACAGCGCTTAAGTTTACAGGACTTTTGGAGTCTTACA 240
QY 842 TATTTGTGACCAAACTTGAAGATGAAACCAAGAAACAGACTTAAACAAATATATCAATGC 901
Db 241 TATTTGTGACCAAACTTGAAGATGAAACCAAGAAACAGACTTAAACAAATATATCAATGC 300
QY 902 AAATGTAATTTTGTGTTTAAAGGCTTGCTTGATGGTCAAGTTATCCCAATGSACA 961
Db 301 AAATGTAATTTTGTGTTTAAAGGCTTGCTTGATGGTCAAGTTATCCCAATGSACA 360
QY 962 CTAAGTTAGACCAACAACAAACCTGATTCGTCTCTTT 1001
Db 361 CTAAGTTAGACCAACAACAAACCTGATTCGTCTCTTT 400

RESULT 7
BG193969/c
LOCUS BG193969 894 bp mRNA linear EST 21-APR-2001
DEFINITION R31111 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG193969
VERSION BG193969.1 GI:13715656
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 894)

REFERENCE

AUTHORS

Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R., Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whittington, J., Lerner, S., Costanzo, D., McElligott, K., Booser, S., Mays, R., Smith, E., Veloso, N., Klika, A., Hess, J., Cothren, K., So, K., Offenbacher, J., Danzig, J., and Ducar, M.

TITLE

Creation of genome-wide protein expression libraries using random activation of gene expression

JOURNAL

MEDLINE

PUBMED

COMMENT

Contact: Scott C. Cain

Athersys, Inc.

3201 Carnegie Ave, Cleveland, OH 44115, USA

Tel: 216 431 9900

Fax: 216 361 9596

Email: scaine@athersys.com

High quality sequence stop: 461

Location/Qualifiers

1..894

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/cell_line="HT1080"

/clone_lib="Athersys RAGE Library"

/note="See 'Creation of Genome-wide Protein Expression Libraries using Random Activation of Gene Expression', Nature Biotechnology, in press. Note that even though the cell type indicated is HT1080, since a random activation method was used, these sequence tags are not necessarily expressed in HT1080 under normal circumstances."

BASE COUNT 284 a 154 c 169 g 283 t 4 others

ORIGIN

Query Match 35.0%; Score 350; DB 10; Length 894;
Best Local Similarity 96.3%; Pred. No. 0.029;
Matches 389; Conservative 0; Mismatches 12; Indels 3; Gaps 3;

QY 599 AGATGAGAAATATTCGATTATTCGACTTCACCTCACTGAATCCTTGAAAAAATAAAGCGCG 658
Db 770 AGATGAGAAATATTCGATTATTCGACTTCACCTCACTGAATCCTTGAAAAAATAAAGCGCG 711
QY 659 TCAGCCCTCAAACTTT-AGAAGCGTACATACGTTGGAACAAGACTTTGGAGATACCCT 717
Db 710 TCAGCCCTCTAACTTTAAGAGCGGACATACGTTGGACCAAGACTTTGGAGATACCCT 651
QY 718 GTTTAAGGAAATACCTTTGTAAAGCTGCAGAACATTTTACTTAAAGAGGAAACACAAGA 777
Db 650 GTTTAAGGAAAT-CTTTGTAAAGCTGCAGAACATTTT-CTTAAAGAGGAAACACAAGA 593
QY 778 TCTCAATGAACGTCATCGGCTACAGAAACAGCGCTTAAGTTTACAGGACTTTTAGAGTCT 837
Db 592 TCTCAATGAACGTCATCGGCTACAGAAACAGCGCTTAAGTTTACAGGACTTTTAGAGTCT 533
QY 838 TACATATTTGTGACCAAACTTGAAGATGAACCAAGAAACAGACTTAAACAAATATACA 897
Db 532 TACATATTTGTGACCAAACTTGAAGATGAACCAAGAAACAGACTTAAACAAATATACA 473
QY 898 ATGCAAAATGTAATTTTGTGTTTAAAGGCTTGCTTGATGGTCAAGTTATCCCAATG 957
Db 472 ATGCAAAATGTAATTTTGTGTTTAAAGGCTTGCTTGATGGTCAAGTTATCCCAATG 413

RESULT 8
BF216343
LOCUS BF216343 728 bp mRNA linear EST 06-NOV-2000

```

DEFINITION 601894376F1 NIH_MGC_57 Homo sapiens cDNA clone IMAGE:4102722 5',
mRNA sequence.
ACCESSION BF216343
VERSION BF216343.1 GI:11109929
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 728)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC).
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: CLONETECH Laboratories, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LJNL)
cDNA Sequencing by: Incyte Genomics, Inc.
cDNA Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LJNL at:
http://image.lni.gov
Plate: LCM976 row: h column: 19
High quality sequence stop: 574.
Location/Qualifiers
1..728
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4102722"
/tissue_type="glioblastoma"
/lab_host="DH10B (T: phage-resistant)"
/clone_lib="NIH_MGC_57"
/notes="Organ: brain; Vector: pDR-LIB (Clontech); Site 1:
SfiI (ggcgccctggcc); Site 2: SfiI (ggccattagggcc);
Double-stranded cDNA was prepared from cell line RNA. 5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGAGCGCGGCGATG-dTAC(BN)-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.55 kb (range 0.9-4.0 kb). 12/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT 248 a 132 c 150 g 198 t
ORIGIN
Query Match 28.6%; Score 286.2; DS 10; Length 728;
Best Local Similarity 92.6%; Pred. No. 1.8;
Matches 377; Conservative 0; Mismatches 23; Indels 7; Gaps 7;

QY 599 AGATGAGAAATATTCGATTATCTGACTTCACTGAACTTGAAGAAATATAAAGCGAGCG 456
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D5 142 AGATGAGAAATATTCGATTATCTGACTTCACTGAACTTGAAGAAATATAAAGCGAG-G 200
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 659 TCAGCGCTCAAACTTTAGAGCGGTACATACGTTTGAACACAGGACTT-TGGAGATACCACT 717
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D5 201 TCAGCGCTCAAA-TTTAGAGCGGTACATACGTTTGAACACAGGACTTGTGGAGATACCACT 259
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 718 -GTTTAAGGAATACCTT-TGTAACCTGCAGAACATTTTACTTTAAAGAGGAAACACAA 775
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D5 260 GGTTTAAGGAATACCTTGTGTAAACCTGCAGAACATTTTACTTTAAAGAGGAAACACAA 319
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 776 GATCTTCAATGAACGTCATCGGCTACAGAAACAGCCCTAAGTTTACAGGACTTTTAGAGT 835
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D5 320 GATCTTCAATGAACGTCATCGGCTACAGAAACAGCCCTAAGTTTACAGGACTTTTAGAGT 379
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 836 CTTACATATTTGTGCACCAAACTTGAAGATGAACCGAAACACAGACTTAAACCAATATA 895
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D5 380 CTTACATATTTGTGCACCAAACTTGAAGATGAACCGAAACACAG-TTAAACCAATATA 438
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 896 CAATGCAATGTAAATTTTGTGTTAAGGCGCTTGCGTTGATGGTCACAGTTATCCCAA 955
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

D5 439 CAATGCAATGTAAATTTTGGTGGTCTTAAGGCGCTTGCTGGGATGGTCACAGTTATCCCAA 498
QY 856 TGGACACTAAG-TTAGAGCACAACAAACCTGATTTCTGCTCTCTTT 1001
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D5 499 TGGACACTAAGTTTAGAGCACAACAAACCTGATTTCTGCTCTCTTT 545
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
BG185583 360 bp mRNA linear EST 21-APR-2001
DEFINITION RST4534 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION BG185583
VERSION BG185583.1 GI:13707270
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 360)
AUTHORS Harrington, J.J., Sherf, B., Rundlett, S., Jackson, P.D., Perry, R.,
Cain, S., Leventhal, C., Thornton, M., Ramachandran, R., Whittington, J.,
Lerner, L., Costanzo, D., McElligott, K., Boorer, S., Mays, R., Smith,
E., Veloso, N., Klika, A., Hess, J., Cothren, K., Lo, K., Offenbacher,
J., Danzig, J. and Ducar, M.
TITLE Creation of genome-wide protein expression libraries using random
activation of gene expression
JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE 21227151
PUBMED 11329013
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave., Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 310.
Location/Qualifiers
1..360
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_lib="HT1080"
/clone="See 'Athersys RAGE Library'"
/notes="See 'Creation of Genome-wide Protein Expression
Libraries Using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT 129 a 61 c 53 g 107 t
ORIGIN
Query Match 27.5%; Score 277.8; DS 10; Length 360;
Best Local Similarity 97.1%; Pred. No. 5;
Matches 304; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

QY 659 GTTGGACAAGGACTTTGGAGATACCACTGTTTAAGSAAATACCTTTGTAACCTGAGCA 748
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D5 15 GTTGGACAAGGACTTTGGAGAT-CCACTGTTGAAGSAAATACCACTGTAACCTGAGCA 73
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 749 ACATTTTACTTAAAGAGGAAACACACAGATCTTCAATGAACGTCATCGGCTACAGAAACA 808
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D5 74 ACATTTTACTTAAAGAGGAAACACACAGATCTTCAATGAACGTCATCGGCTACAGAAACA 133
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 809 GCCTAAGTTTACAGGACTTTTGTAGAGTCTTACATATTTGTGCACCAAACTTGAAGATGAA 868
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D5 134 GCCT-AGTTTACAGGACTTTTGTAGAGTCTTACATATTTGTGCACCAAACTTGAAGATGAA 192
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 869 CCAGAAACAGACTTAAACAAATATACATGCAATGCAATTTTGTGTTTAAAGGCC 928
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
D5 193 CCAGAAACAGACTTAAACAAATATACAGTGCAATGCAATTTTGTGTTTAAAGGCC 252
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 929 TTGCCTTGATGGTCACAGTTATCCCAATGGACACTAGTTAGAGCACAACAAACCTGAT 988
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Db 253 TTGCTTGTGATGGTACAGTTATCCCAATGGACACTAGGTAGAGACACAAACCTGAT 312
QY 989 TCTGGTCTCTTT 1001
Db 313 TCTGGTCTCTTT 325

RESULT 10
BE17274C/C
LOCUS BE172740 460 bp mRNA linear EST 21-APR-2000
DEFINITION YR0-HT0559-120400-010-009 HT0559 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE172740
VERSION BE172740.1 GI:8635466
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 460)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.E.,
Nagal,M.A., da Silva,W.C., Zagac,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Balaz,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,
M.J., Soares,F., Brentani,R.R., Reis,H.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (21), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=ht2=MR0-HT0559-120
400-010-a09&t3=2000-04-12&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 6
High quality sequence stop: 460.
Location/Qualifiers
1..460
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="HT0559"
/note="Organ: head/neck; Vector: puc18; Site: 1, Small
Site: 2, Small; A mini-library was made by cloning products
derived from CRESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
117 a 112 c 84 g 147 b

BASE COUNT 117 a 112 c 84 g 147 b
ORIGIN

Query Match 24.4%; Score 244.6; DB 10; Length 460;
Best Local Similarity 74.6%; Pred. No. 33;
Matches 326; Conservative 0; Mismatches 99; Indels 12; Gaps 1;

QY 5 TAAAAATGATACCTTAGGCTGGGCAAGTGGCTACGCTGTGTATCCAGCAGCTTGGG 64
Db 460 TAATAATAATAAAAAAGGCCAGGTGGGTGGTTCACACCTGTATCCAGCAGCTTGGG 401
QY 65 AGGCCAAGTGGCGGATACCTGAGCTAGGAGTTGAGACCAACCTGGCCATCTGGC 124

Db 400 AGGCCACGACAGGTGGATCATGAGTCAGGAGTTCGAGACCAGCCTGGCCAAACGGTGT 341
QY 125 GAAACCCGTGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGAGGATGTGCCTGTAA 184
Db 340 GAAACCCGTGCTCTCTACTAAAAACACAAAAATTAGCTAGGCGTGGGGGGGAGGCTGTAA 281
QY 185 TCCAGCTACTCGGGAGGCTGAGACAGAGAATTTGTTGAACCTTGGAGGGGAGGTTC 244
Db 280 TCCAGCTACTAGGAGGCTTGAGGCAGGAGATCACTGAATCTGGAAAGCAGAGGTTC 221
QY 245 ATATCTAGTGTGAAATTTGTGATTC-----TTTTTCTCTTTGTGTGATTT 292
Db 210 AGTGAGCGGAGGTTCACACCATTTGCACCTCCAGCCTGGGCGACAAAGATCTGTCTC 161
QY 293 TTGAACTTTTCTATTAATGATTGTGTTTGTGTTTATATATGGAAAAATATTATGCTTC 352
Db 140 AAAAAATATATATATAAATGATTGTGTTTGTGTTTATATATGGAAAAATATTATGCTTC 101
QY 353 AAATGTTAATACCTATGAACCTAAACACAAAGTAATAAATATATTAGTATAGCAATTATTA 412
Db 100 AAATGTTAATACCTATGAACCTAGACACAAAGTAATAAATATATTAGTATAGCAATTATTA 41
QY 413 AGGTTTCTTGTGTAGCA 429
Db 40 AGGTTTCTTGTGTAGCA 24

RESULT 11
RG208384
LOCUS RG208384 239 bp mRNA linear EST 21-APR-2001
DEFINITION RST27882 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
ACCESSION RG208384
VERSION RG208384.1 GI:13730371
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 239)
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,C.,
Lerner,L., Costanzo,D., McEligott,K., Boozer,S., Mays,R., Smith,
E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher,
J., Danzig,J. and Ducar,M.
TITLE Creation of genome-wide protein expression libraries using random
activation of gene expression
JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE 21227151
PUBMED 11329013
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 189.
Location/Qualifiers
1..239
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_line="HT1080"
/clone_lib="Athersys RAGE Library"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
89 a 46 c 42 g 63 t

BASE COUNT 89 a 46 c 42 g 63 t
ORIGIN

Query Match 21.8%; Score 218.2; DB 10; Length 239;


```

QY 6 AAAAAATGATACCTTTAGGCTGGGCAAAAGTGGCTCAGCTGTGTATATCCAGGACACTTTGGGA 65
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 506 AAAAAATGCAAAAATTAGGCTGGGCGGAGTGGCCCGAGGCTGTAAATCCAGGACACTTTGGGA 447
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 56 GGGCCAGGTGGGCGGATCACCTTGAGGTTAGGAGTTAGGACCAAGATGGCCATATGSCG 125
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 446 GGGCCAGGTGGGTTGGATCACCTTAAGTTCAGGAGTTAGGACCAAGATGGCCATATGSCG 387
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 126 AAACCCCTGTCTCTACTATAAATACAAAATTTAGCTGGGCACTGGAGGATGTGCCTGTAAAT 185
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 386 AAACCCCACTCTCTACTATAAATACAAAATTTAGCTGGGCACTGGAGGATGTGCCTGTAAAT 327
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 186 CCCAGCTACTCGGGAGGCTGAGACAGAGAGANTTCCTTGACACTTGGAGGGGAGGTTCGA 245
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 326 CCCAGCTACTCGGGAGGCTGAGACAGAGAGANTTCCTTGACACTTGGAGGGGAGGTTCGA 267
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
CA946732/c
LOCUS CA946732 515 bp mRNA linear EST 21 DEC-2002
DEFINITION is09h12.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6364654 3',
        mRNA sequence.
ACCESSION CA946732
VERSION CA946732.1 GI:27435609
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 515)
REFERENCE Lemishka, J., Scearce, N., Brestolli, G., Gradwohl, G., Clifton, S.,
            Hillier, L., Marra, M., Pape, D., Wyllie, T., Martin, J., Blistain, A.,
            Schwartz, A., Theising, P., Ritter, E., Ronko, J., Bennett, S., Cardenas
            M., Gibbons, K., McCann, R., Cole, R., Tsagaris, R., Williams, T.,
            Jackson, Y., and Bowers, Y.
            Endocrine Pancreas Consortium
            Unpublished
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@im.wustl.edu)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 451.
            Location/Qualifiers
                1..515
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:6364654"
                /tissue_type="Purified pancreatic islet"
                /lab_host="DH1CB"
                /clone_lib="HR85 islet"
                /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                Not1; Site_2: XhoI; cDNA made by oligo-dT priming.
                Size-selected on agarose gel. Average insert size ~1kb. 5'
                XhoI site was destroyed after directional cloning.
                Amplified once. Contact information: Hiroshi Inoue, MD,
                Metabolism Div. (Alan Permutt Lab), Washington University
                School of Medicine, Box 8127, 660 South Euclid Ave., St.
                Louis, MO 63110, E-mail: hinoue@im.wustl.edu, Tel:
                314-362-1916, Fax: 314-747-2692."
BASE COUNT 94 a 154 c 114 g 153 t
ORIGIN

```

```

Query Match 12.9% Score 188.8; DB 14; Length 515;
Best Local Similarity 85.7%; Pred. No. 9.8e+02;
Matches 258; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 5 AAAAAATGATACCTTTAGGCTGGGCAAAAGTGGCTCAGCTGTGTATATCCAGGACACTTTGGGA 65
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 297 AAAAAATACAAAATTAGGCTGGGCGGAGTGGCCCGAGGCTGTAAATCCAGGACACTTTGGGA 238
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 56 GGGCCAGGTGGGCGGATCACCTTGAGGTTAGGAGTTAGGACCAAGATGGCCATATGSCG 125
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 237 GGGCCAGGTGGGTTGGATCACCTTAAGTTCAGGAGTTAGGACCAAGATGGCCATATGSCG 178
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 126 AAACCCCTGTCTCTACTATAAATACAAAATTTAGCTGGGCACTGGAGGATGTGCCTGTAAAT 185
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 177 AAACCCCACTCTCTACTATAAATACAAAATTTAGCTGGGCACTGGAGGATGTGCCTGTAAAT 118
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 186 CCCAGCTACTCGGGAGGCTGAGACAGAGAGANTTCCTTGACACTTGGAGGGGAGGTTCGA 245
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 117 CCCAGCTACTCGGGAGGCTGAGACAGAGAGANTTCCTTGACACTTGGAGGGGAGGTTCGA 58
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 14
AI339725/c
LOCUS AI339725 446 bp -RNA linear EST 13-FEB-1999
DEFINITION J942e08.x1 Soares_NHMPU_S1 Homo sapiens cDNA clone IMAGE:1935206
        1' similar to contains Alu repetitive element; contains element PTR5
        repetitive element 1; mRNA sequence.
ACCESSION AI339725
VERSION AI339725.1 GI:4076552
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 446)
REFERENCE NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished
            Contact: Robert Strausberg, Ph.D.
            Email: cgaps-r@mail.nih.gov
            This clone is available royalty-free through LINDA; contact the
            IMAGE Consortium (info@image.lni.gov) for further information.
            Insert Length: 2084 Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 446.
            Location/Qualifiers
                1..446
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="IMAGE:1935206"
                /tissue_type="Pooled human melanocyte, fetal heart, and
                pregnant uterus"
                /lab_host="DH1CB"
                /clone_lib="Soares_NHMPU_S1"
                /note="Organ: mixed (see below); Vector: pTZ193D-Pac
                (Barracina) with a modified polylinker; Site 1: Not 1;
                Site 2: Eco RI; Equal amounts of plasmid DNA from three
                normalized libraries (melanocyte 2NbHM, pregnant uterus
                NBHPU, and fetal heart NBH19W) were mixed, and ss circles
                were made in vitro. Following HAP purification, this DNA
                was used as tracer in a subtractive hybridization
                reaction. The driver was PCR-amplified cDNAs from pools of
                5,000 clones made from the same 3 libraries. The pools
                consisted of I.M.A.G.E. clones 260232-265223,
                340488-345479, and 484488-489479."
                340488-345479, and 484488-489479."
BASE COUNT 120 a 117 c 99 g 110 t
ORIGIN

```

Query Match 18.7% Score 187.4; DB 9; Length 446;
Best Local Similarity 85.3%; Pred. No. 1.2e+03;

Matches	209;	Conservative	3;	Mismatches	36;	Indels	0;	Gaps	0;
QY	1	AAATATAAAATGCATACCTTTAGGCTGGGCAAGTGGCTACGCTGTGTATATCCAGCACTT	60						
Db	300	AAATTATTAGCAATTAACTCGGCTGGGCGGCTGGCTACGCTGTGTATATCCAGCACTT	241						
QY	61	TGGGAGGCCAAGGTGGGCGGATCACCTGAGGTTGGAGTTGAGGACCAACCTGGCCATCA	120						
Db	240	TGGGAAGCCAGGTGGGCGGATCACCTTCAGGTTAGGAGTTGAGGACCAACCTGGCCATCA	181						
QY	121	TGGCGAACCCTGTCTCTACTATAAATAACAAAATTAGCTGGGCATGAGGCACTGTGCCT	180						
Db	180	TGGCGAACCCTGTCTCTACTATAAATAACAAAATTAGCTGGGCATGAGGCACTGTGCCT	121						
QY	181	GTANTCCAGCTACTCGGGAGGCTGAGACAGAGAATTGCTTGAACCTGGAGGGGAGG	240						
Db	120	GTANTCCAGCTACTCGGGAGGCTGAGGACAGAGAATCGATTAAACCTGGAGGGGGAGA	61						
QY	241	TTGCA	245						
Db	60	TTGCA	56						

Search completed: October 24, 2003, 22:51:13
Job time : 2095.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

GM-nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:43:05 : Search time 1284.09 seconds
(without alignments):
14809.112 Million cell updates/sec

Title: US-09-830-902-1_COPY_10000_10500

Perfect score: 501

Sequence: 1 ggttcctcgctggctctggg.....gggtggggggggggggggc 501

Scoring table: IDENTITY NUC

Gapop 100 , Gapext 1.0

Searched: 2888711 seqs, 2045461386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb_ba:

2: gb_htg:

3: gb_in:

4: gb_em:

5: gb_ov:

6: gb_pat:

7: gb_ph:

8: gb_pl:

9: gb_pr:

10: gb_rc:

11: gb_sts:

12: gb_sy:

13: gb_un:

14: gb_vit:

15: em_ba:

16: em_fun:

17: em_hum:

18: em_in:

19: em_mu:

20: em_om:

21: em_or:

22: em_ov:

23: em_pat:

24: em_ph:

25: em_pl:

26: em_rc:

27: em_sts:

28: em_un:

29: em_vit:

30: em_htg_hum:

31: em_htg_invi:

32: em_htg_other:

33: em_htg_mus:

34: em_htg_pln:

35: em_htg_rod:

36: em_htg_mam:

37: em_htg_vrt:

38: em_sy:

39: em_htgc_hum:

40: em_htgc_mus:

41: em_htgc_other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	501	100.0	101584	9	CNS01DS5	AL121655 BAC seque
2	501	100.0	110000	6	AXC93471	AXC93471 Sequence
3	501	100.0	110000	9	HSA246003	AC246003 Homo sapi
4	501	100.0	15943	9	ACC12364	ACC12364 Homo sapi
5	492	94.2	3263	6	AXC93472	AXC93472 Sequence
6	472	94.2	3263	9	HSA246001	AC246001 Homo sapi
7	472	94.2	5120	9	ABC29006	AB029006 Homo sapi
8	442	88.2	185281	2	ACC11232	AC011232 Homo sapi
9	348.8	69.6	2152	6	AXC93576	AXC93576 Sequence
10	348.2	69.5	969	9	HSA325510	AJ325510 Homo sapi
11	346	69.1	2116	10	BC046286	BC046286 Mus muscu
12	274	54.7	71793	2	AC101733	AC101733 Mus muscu
13	247.4	49.4	721	9	HSA342191	AC342191 Homo sapi
14	213.4	42.6	712	9	HSA342199	AJ342199 Homo sapi
15	163.8	33.1	697	9	HSA342196	AJ342196 Homo sapi
16	73	14.6	1689	6	AXC93542	AXC93542 Sequence
17	73	14.6	1689	10	MMJ246002	AJ246002 Mus muscu
18	64	12.8	125020	9	AF429315	AF429315 Homo sapi
19	59.4	11.9	161922	2	AC144461	AC144461 Mus muscu
20	58.6	11.7	133575	8	CNS08C9N	AL732646 Oryza sat
21	58.2	11.6	37159	2	AC100259	AC100259 Mus muscu
22	57.2	11.4	65942	2	AC036124	AC036124 Homo sapi
23	57	11.4	200342	9	CNS01RG3	AL157736 Human chr
24	56.6	11.3	7218	6	166494	166494 Sequence 14
25	55.8	11.1	135119	2	AC011578	AC011578 Homo sapi
26	55.4	11.1	154728	9	AC018730	AC018730 Homo sapi
27	54.8	10.9	131602	2	AC034263	AC034263 Homo sapi
28	54.4	10.9	80133	2	AC021347	AC021347 Homo sapi
29	54.2	10.8	63082	2	AC022663	AC022663 Homo sapi
30	53.8	10.7	159028	2	AC142003	AC142003 Rattus no
31	53.4	10.7	300695	2	AC079431	AC079431 Mus muscu
32	53.2	10.6	87283	2	AC022856	AC022856 Homo sapi
33	53.2	10.6	115402	2	AC027283	AC027283 Homo sapi
34	53	10.6	159851	2	AP003571	AP003571 Oryza sat
35	53	10.6	171574	2	AC012300	AC012300 Homo sapi
36	53	10.6	219952	2	AC084804	AC084804 Mus muscu
37	52.8	10.5	288220	9	AC009107	AC009107 Homo sapi
38	52.6	10.5	40343	2	AC099831	AC099831 Homo sapi
39	52.4	10.5	2837	9	HSCGG16P5	AJ001218 Homo sapi
40	52.4	10.5	139087	2	AP004797	AP004797 Oryza sat
41	52.4	10.5	141534	2	AP003685	AP003685 Oryza sat
42	52.4	10.5	155419	2	AP005840	AP005840 Oryza sat
43	52.4	10.5	189865	2	AP003766	AP003766 Oryza sat
44	52.2	10.4	64211	2	AC091110	AC091110 Homo sapi
45	52.2	10.4	133790	2	AC119292	AC119292 Oryza sat

ALIGNMENTS

RESULT :
CNS01DS5
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

CNS01DS5
101584 bp DNA linear
BAC sequence from the SPG4 candidate region at 2p21-2p22 BAC 316P14
of library Cifg_978_SKB from chromosome 2 of Homo sapiens (Human).
AL121655
AL121655.1 GI:6002388
SPG4 genomic DNA interval.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 101584)
Hazar, C., Fokkrecht, N., Mavel, D., Paternotte, C., Samson, D.,
Artiguenave, F., Davoine, C.S., Cruaud, C., Durr, A., Wincker, P.,

Brottier, P., Cattolico, L., Bayle, V., Burgunder, G.M.,
 Prud'Homme, J.F., Brice, A., Fontaine, B., Heilig, R. and
 Weissenbach, J.

Spastin, a novel AAA protein, is altered in the most frequent form
 of autosomal dominant spastic paraplegia
 Nat. Genet. (1999) in press

REFERENCE 2 (bases 1 to 101584)

AUTHORS

JOURNAL

Direct Submission

Submitted (19-APR-2002) Genoscope - Centre National de Sequencage

BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr)

- Web : www.genoscope.cns.fr

Location/Qualifiers

1. 101584

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="2"

/clone="336P.4"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

/clone_lib="CITS 978 SKB"

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

Patent: WO 01/0198-A 1 15-MAR-2002

Centre National de Recherche Scientifique (CNRS) (FR)

Location/Qualifiers

1. 110000

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

9932. 10471

10472. 33719

33719. 33605

33606. 35748

35749. 35832

35833. 45022

45023. 45118

45119. 60863

60864. 61051

61052. 61927

61928. 62061

62062. 62344

62345. 62438

62439. 73173

73174. 73248

73249. 74633

74634. 74705

74706. 82788

82789. 82864

82865. 83102

83103. 83194

83195. 83334

83335. 83414

83415. 88129

88130. 89172

89173. 89561

89562. 89641

89642. 91162

91163. 91233

91234. 93443

93444. 93484

93485. 100599

100600. 102009

BASE COUNT 30622 a 21640 c 22817 g 34921 t

ORIGIN

Query Match 100.0%; Score 501; DB 6; Length 110000;

Best Local Similarity 100.0%; Pred. No. 1.4e-74;

Mismatches 0; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2

AX093471

LOCUS

DEFINITION

Sequence 1 from Patent WO0118198.

ACCESSION

AX093471

VERSION

AX093471.1 GI:13509911

KEYWORDS

Homo sapiens (human)

SOURCE

Homo sapiens

ORGANISM

Homo sapiens

AX093471 110000 bp DNA linear PAT 30-MAR-2001

Sequence 1 from Patent WO0118198.

AX093471

AX093471.1 GI:13509911

Homo sapiens (human)

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

QY	301	TTCTCCGCGCCCTCATCGCAGCCACAGAGGAGCTTCGGGGCGCGCCAGCACCTGCCTCG	363
DB	10360	TTCTCCGCGCCCTCATCGCAGCCACAGAGGAGCTTCGGGGCGCGCCAGCACCTGCCTCG	10359
QY	361	GCCTCGCCCGCGCGCGCTGCGCGCGCGAGCGCGGCTCCAGTCTCCACAAA	423
DB	10360	GCCTCGCCCGCGCGCGCTGCGCGCGCGAGCGCGGCTCCAGTCTCCACAAA	10419
QY	421	CAGCGCTTCGAGTACATCTCCATTCGCGCTCGCGATCGGATCGAAGGTAAGTAG	483
DB	10420	CAGCGCTTCGAGTACATCTCCATTCGCGCTCGCGATCGGATCGAAGGTAAGTAG	10479
QY	481	GGGGCTGGGGAGGGGGGGG	501
DB	10480	GGGGCTGGGGAGGGGGGGG	10500
RESULT 3			
HSA246003			
LOCUS			
DEFINITION Homo sapiens Spast gene for spastin protein.			
ACCESSION AC246003			
VERSION AC246003.1 GI:6273492			
KEYWORDS Spast gene; spastin protein; SPG4-linked hereditary spastic paraplegia.			
SOURCE Homo sapiens (human)			
ORGANISM Homo sapiens			
REFERENCE 1			
AUTHORS Hazan, J., Fonknechten, N., Mavel, D., Paternotte, C., Samson, D., Artiguerave, F., Davoine, C.S., Cruaud, C., Durr, A., Wincker, P., Brothier, P., Catolico, L., Barbe, V., Burgunder, J.M., Prud'homme, J.F., Brice, A., Fontaine, B., Heilig, R. and Weissenbach, J.			
TITLE Spastin, a new AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia			
JOURNAL Nat. Genet. 23 (3), 296-303 (1999)			
MEDLINE 20055425			
PubMed 10610178			
REFERENCE 2			
AUTHORS Fonknechten, N., Mavel, D., Byrne, P., Davoine, C., Cruaud, C., Roentsch, D., Samson, D., Coutinho, P., Hutchinson, M., McKusick, P., Burgunder, J., Tattiagione, A., Heinzel, O., Fekli, E., Deifel, T., Parfrey, N., Brice, A., Fontaine, B., Prud'homme, J., Weissenbach, J., Durr, A. and Hazan, J.			
TITLE Spectrum of SPG4 mutations in autosomal dominant spastic paraplegia			
JOURNAL Hum. Mol. Genet. 9 (4), 637-644 (2000)			
MEDLINE 20164302			
PubMed 10699187			
REFERENCE 3 (bases 1 to 110000)			
AUTHORS Genoscope.			
TITLE Direct Submission			
JOURNAL Submitted (17-JUN-1999) Genoscope, Genoscope - Centre National de Sequencage, BP 191, EVRY 91006, FRANCE			
COMMENT E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr. The sequence is the result of the assembly of 2 BAC clones: R-336p14 and 563N4, respectively from RPC1-11 and CITE_978_SKH library.			
FEATURES			
SOURCE			
1. 110000			
/organism="Homo sapiens"			
/mol_type="genomic DNA"			
/db_xref="taxon:9606"			
/chromosome="2"			
/map="2p21"			
/clone="R-336P14"			
/clone_id="RPC1-11"			
/note="BAC R-336P14 contains Spast gene: exon 1 to exon 3, BAC 563N4 contains Spast gene: exon 4 to exon 17"			
9932..102009			
gene			
/gene="Spast"			

mRNA	9932..102009	/gene="Spast"
exon	/product="spastin protein"	
	9932..10471	/db_xref="GI:6273492"
	/gene="Spast"	
	/number=1	
	9932..10056	/db_xref="GI:6273492"
	/gene="Spast"	
	join(1057..10471,33719..33805,35749..35832,45023..45118,	
	62864..62051,61928..62061,62345..62438,73174..73248,	
	74634..74705,82789..82863,83103..83194,83335..83414,	
	88130..88172,89582..89641,91163..91233,93444..93484,	
	100600..100722;	
	/gene="Spast"	
	/codon_start=1	
	/product="spastin protein"	
	/protein_id="CA860238.1"	
	/db_xref="GI:6273492"	
	/translation="YNSPGGRGKKKSGASNPVPRPPPPCLAPAPPAAGPAPPES	
	PHKPNLYPSYPLFVGFALRLVAFHJLGLFWLQRFRRALMAKRSSGAAPAPASA	
	SAPAPVPGGEAEVRVFRHQAPEYISIALRIDEDERAGKEQAVEWYKKGIELEKGI	
	AVITGCGGCCERARRLCAKMTNLVMAKDRLOLLEKYPVLPPSKQTDVYNDSTNL	
	ACRNGHLOSEGAVPKRDPJTHTSNSTRSKTMKTGAGLSGHHRAPSYSGLSVVS	
	GVKQSGRPATTHKGTPKTRTNKPSSTTTATKKKELKFNVESNLINNEIVD	
	NGTAVKFDIAGODLAKALQELVILPSIRPELFTGLPAPARGLLIFGPPGNGKTMIA	
	KAVAAENATFFNISAASLTSKYVGESEKIVRAJFAVARELOPSTTFIDEVDSLQER	
	REGHDASRLKTEFLIFDGVQSAGDRLVMGATNRQELDEAVLRPIKRVVSL	
	PNEETRJLLKLLCKQSPLTCKELAJLARMTDGYSGSLTALAKDAALGPIRELKP	
	EQVKMMSASEMRNIRLSPTESLKKIKRVSVPQTLQAVIRWKKDPGDTTV"	
intron	10472..33718	/gene="Spast"
	/number=1	
	33719..33805	/number=2
exon	/gene="Spast"	
	/number=2	
intron	33806..35748	/gene="Spast"
	/number=2	
exon	35749..35832	/gene="Spast"
	/number=3	
intron	35833..45022	/gene="Spast"
	/number=3	
exon	45023..45118	/gene="Spast"
	/number=4	
intron	45119..60863	/gene="Spast"
	/number=4	
exon	60864..61051	/gene="Spast"
	/number=5	
intron	61052..61927	/gene="Spast"
	/number=5	
exon	61928..62061	/gene="Spast"
	/number=6	
intron	62062..62344	/gene="Spast"
	/number=6	
exon	62345..62438	/gene="Spast"
	/number=7	
intron	62439..73173	/gene="Spast"
	/number=7	
exon	73174..73248	/gene="Spast"
	/number=8	
intron	73249..74633	/gene="Spast"

```
/number=8
74634..74704
/gene="Spast"
/number=9
74706..74788
/gene="Spast"
/number=9
82789..82864
/gene="Spast"
/number=10
82865..83102
/gene="Spast"
/number=10
83103..83194
/gene="Spast"
/number=11
83195..83334
/gene="Spast"
/number=11
83335..83414
/gene="Spast"
/number=12
83415..88129
/gene="Spast"
/number=12
88130..88172
/gene="Spast"
/number=13
88173..89561
/gene="Spast"
/number=13
89562..89641
/gene="Spast"
/number=14
89642..91162
/gene="Spast"
/number=14
91163..91233
/gene="Spast"
/number=15
91234..93443
/gene="Spast"
/number=15
93444..93484
/gene="Spast"
/number=16
93485..100509
/gene="Spast"
/number=16
100510..102009
/gene="Spast"
/number=17
102010..102009
/gene="Spast"
101973..101978
/gene="Spast"
BASE COUNT 30622 a 21640 c 22817 g 34921 t
ORIGIN
Query Match: 100.0%; Score 501; DB 9; Length 110000;
Best Local Similarity 100.0%; Pred. No. 1.4e-74;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGTTCGGTCGGTTCGGGAGCGGGTTATGCGCGCGCGCGGAGTGTGAGCTGTGAATG 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10000 GGTTCGGTCGGTTCGGGAGCGGGTTATGCGCGCGCGCGGAGTGTGAGCTGTGAATG 10059
QY 61 AATTCTCCGGTGGACGAGGGGAAGAAAGGCTCCGGCGCGCGGAGCAACCGCGTGCCT 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10060 AATTCTCCGGTGGACGAGGGGAAGAAAGGCTCCGGCGCGCGGAGCAACCGCGTGCCT 10119
QY 121 CCCAGGCTCCGCCCGCTTGCCTGGCCCGCCCGCTCCCGCGCGCGGCGCGCTCCG 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 10120 CCCAGGCTCCGCCCGCTTGCCTGGCCCGCCCGCTCCCGCGCGCGGCGCGCTCCG 10179
QY 181 CCCAGTCCGGGCATAGCGAACCCTGTACTATTCTCTACCGCGCTGTGTAGGCTTC 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10180 CCCAGTCCGGGCATAGCGAACCCTGTACTATTCTCTACCGCGCTGTGTAGGCTTC 10239
QY 241 GCGCTGCTGCGTTGGTCCCTTCCACCTGGGGCTCCCTTCTGCTGTGGCTCTGCCAGCGC 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10240 GCGCTGCTGCGTTGGTCCCTTCCACCTGGGGCTCCCTTCTGCTGTGGCTCTGCCAGCGC 10299
QY 301 TCTCCCGCGCCCTCATGGCAACCAAGAGGAGCTCCGGGCGCGGCCAGCACTGCCTCG 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10300 TCTCCCGCGCCCTCATGGCAACCAAGAGGAGCTCCGGGCGCGGCCAGCACTGCCTCG 10359
QY 361 GCGTCGGCCCGCGCGCGGTGCGGGCGCGGAGCGCGGAGCGCTCCGAGCTTCCACAAA 420
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10360 GCGTCGGCCCGCGCGCGGTGCGGGCGCGGAGCGCGGAGCGCTTCCGAGCTTCCACAAA 10419
QY 421 CAGGCTTCGAGTACATCTCCATTGCCCTGGCATCGATGAGGATGAGAAAGTAACTAG 480
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10420 CAGGCTTCGAGTACATCTCCATTGCCCTGGCATCGATGAGGATGAGAAAGTAACTAG 10479
QY 481 GCGCTGGGGGAGGGGGCGGC 501
||||| ||||| ||||| ||||| |||||
Db 10480 GCGCTGGGGGAGGGGGCGGC 10500

RESULT 4
AC012364 155943 bp DNA linear PRI 09-JAN-2002
LOCUS Homo sapiens BAC clone RP11-444D15 from 2, complete sequence.
DEFINITION AC012364
ACCESSION AC012364
VERSION AC012364.7 GI:14589736
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 155943)
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11): 1097-1108 (1998).
MEDLINE 99063792
PubMed 9847074
REFERENCE 2 (bases 1 to 155943)
AUTHORS Ali, N. and Abbott, A.
TITLE The sequence of Homo sapiens BAC clone RP11-444D15
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 155943)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (25-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 155943)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (03-JULY-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 155943)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jul 3, 2001 this sequence version replaced g11602953.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0444D15
```


MEDLINE 20164302
PubMed 20699187
REFERENCE 3 (bases 1 to 3263)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1999) Genoscope, Genoscope - Centre National de
Sequencage, BP 191, EVRY 91006, FRANCE
COMMENT E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr.
FEATURES Location/Qualifiers
source
1..3263
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2p21"
1..3263
/gene="Spast"
126..1976
/gene="Spast"
/codon_start=1
/product="spastin protein"
/protein_id="CA86014.1"
/db_xref="GI:6273491"
/translation="MNSPGGRGKKKSGGASNPVPRPPPCLLPAPPAAGPAPPES
PHKRLNYFSYPLFVGFALLRLVAFHGLGLFWLQCFRFRALMAKRSAGAPAPASA
SAPAPVGGGEAERVVPHKQAFYISIALRIDEDKAGQKEQAVENYKKGIELEKGI
AVITGQGEQECERARRLOAKMTNLVMAKDRLOJLESQVAPVKKDPJTHISNLSRPSK
ACRNCHQSESGAVPKXKDPJTHISNLSRPSKTYKTSAGLSHRRAPSYSLSMVS
GVKQSGAPATTHKGFPTKTRNKPSPTTATRKKKDLKRNVDSEDLNINVEIVD
NGTAVRDEDDIAGQDLAKQALQEIIVILPSLRPELFGRLAPARGLLFGPPGNGKMLA
KAAVESNATFFNISASLTSKYVGESEKLVRAFAVARELQSEIIFIDEVDSLLCER
REGEDASRLKTEPLIEFDGVSAGDERVLVAGATNRPQSEDAVLRFPIKRVVYSL
PNEETRLLLKLLCKQGSPLTQKEAQLARMTDQYSGSDTLAKIAALGPIRELFK
ECVKNYSASEMRNIRLSDFTESLKKIKRVSYPOTLEAVIRWNKDFGCTTV"
polyA_signal 3227..3232
/gene="Spast"

gene
CDS
polyA_signal 3227..3232
/gene="Spast"
BASE COUNT 956 a 664 c 703 g 940 t
ORIGIN
Query Match 94.2%; Score 472; DB 9; Length 3263;
Best Local Similarity 100.0%; Pred. No. 1.7e-69;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTTCGGTGGTCTGCGGAGGCGGGTTATGGCGGCGGCGGAGTGAGAGTGTGAATG 60
|||
Db 69 GGTTCGGTGGTCTGCGGAGGCGGGTTATGGCGGCGGCGGAGTGAGAGTGTGAATG 129
|||
QY 61 AATTCTCCGGTGGAGGAGGAGAAAGGCTCGCGGCGGCGGAGCAACCGGTGCCT 120
|||
Db 129 AATTCTCCGGTGGAGGAGGAGAAAGGCTCGCGGCGGCGGAGCAACCGGTGCCT 188
|||
QY 121 CCCAGGCTCGGCGGCTTGCCTGGCGGCGGCGGCTCGCGGCGGCGGCGGCGGCTCGG 180
|||
Db 189 CCCAGGCTCGGCGGCTTGCCTGGCGGCGGCGGCTCGCGGCGGCGGCGGCGGCTCGG 248
|||
QY 181 CCGGAGTGGCGGCAAGCGGGAACCTGTACTATTTCTCTACCGGCTGTTGTAGGCTTC 240
|||
Db 249 CCGGAGTGGCGGCAAGCGGGAACCTGTACTATTTCTCTACCGGCTGTTGTAGGCTTC 308
|||
QY 241 GCGTGTGCGGTTTGGTGGCTTCCACCTGGGCTCTCTTCTGTGGCTTGGCAGCGC 300
|||
Db 309 GCGTGTGCGGTTTGGTGGCTTCCACCTGGGCTCTCTTCTGTGGCTTGGCAGCGC 368
|||
QY 301 TTCTCCGCGGCTTATGGCAGCCAGAGAGGAGTCCGGGCGGCGGCGGCGGCTGCTCG 360
|||
Db 369 TTCTCCGCGGCTTATGGCAGCCAGAGAGGAGTCCGGGCGGCGGCGGCGGCTGCTCG 429
|||
QY 361 GCGTGGGCGGCGGCGGCTGCGGCGGCGGAGGCGGCGGCTGCGGAGTCTTCCACAA 420
|||
Db 429 GCGTGGGCGGCGGCGGCTGCGGCGGCGGAGGCGGCGGCTGCGGAGTCTTCCACAA 488
|||
QY 421 CAGGCGCTCGAGTACATCTCCATTGCGGCTGGCATGGATGAGGATGAGAAAG 472
|||

Db 489 CAGGCGCTCGAGTACATCTCCATTGCGGCTGGCATGGATGAGGATGAGAAAG 540
RESULT 7
AB029006
LOCUS AB029006 5120 bp mRNA linear PRI 04-AUG-1999
DEFINITION Homo sapiens mRNA for KIAA1083 protein, complete cds.
ACCESSION AB029006
VERSION AB029006.1 GI:5689502
KEYWORDS
SCJRCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Kikuno,R., Nagase,T., Ishikawa,K., Hiroseawa,M., Miyajima,N.,
Tanaka,A., Kotani,H., Nomura,N. and Chara,C.
Prediction of the coding sequences of unidentified human genes.
XIV. The complete sequences of 100 new cDNA clones from brain which
code for large proteins in vitro
DNA Res. 6 (3), 197-205 (1999)
99397452
10470851
2 (bases 1 to 5120)
Chara,C., Nagase,T. and Kikuno,R.
Direct Submission
Submitted (17-JUN-1999) Osamu Chara, Kazusa DNA Research Institute,
Laboratory of DNA Technology, Yana 1532-3, Kisarazu, Chiba
232-0812, Japan (E-mail:cdnainfo@kazusa.or.jp, Tel:81-438-52-3913,
Fax:+81-438-52-3914)
Location/Qualifiers
1..5120
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="2"
/clone="hJ07423"
/tissue_type="brain"
/clone_lib="pBluescript SK plus"
1..5120
/gene="KIAA1083"
222..1976
/gene="KIAA1083"
/codon_start=1
/product="KIAA1083 protein"
/protein_id="BAA83035.1"
/db_xref="GI:5689503"
/translation="MNSPGGRGKKKSGGASNPVPRPPPCLLPAPPAAGPAPPES
PHKRLNYFSYPLFVGFALLRLVAFHGLGLFWLQCFRFRALMAKRSAGAPAPASA
SAPAPVGGGEAERVVPHKQAFYISIALRIDEDKAGQKEQAVENYKKGIELEKGI
AVITGQGEQECERARRLOAKMTNLVMAKDRLOJLESQVAPVKKDPJTHISNLSRPSK
ACRNCHQSESGAVPKXKDPJTHISNLSRPSKTYKTSAGLSHRRAPSYSLSMVS
GVKQSGAPATTHKGFPTKTRNKPSPTTATRKKKDLKRNVDSEDLNINVEIVD
NGTAVRDEDDIAGQDLAKQALQEIIVILPSLRPELFGRLAPARGLLFGPPGNGKMLA
KAAVESNATFFNISASLTSKYVGESEKLVRAFAVARELQSEIIFIDEVDSLLCER
REGEDASRLKTEPLIEFDGVSAGDERVLVAGATNRPQSEDAVLRFPIKRVVYSL
PNEETRLLLKLLCKQGSPLTQKEAQLARMTDQYSGSDTLAKIAALGPIRELFK
ECVKNYSASEMRNIRLSDFTESLKKIKRVSYPOTLEAVIRWNKDFGCTTV"

FEATURES
source
1..5120
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="2"
/clone="hJ07423"
/tissue_type="brain"
/clone_lib="pBluescript SK plus"
1..5120
/gene="KIAA1083"
222..1976
/gene="KIAA1083"
/codon_start=1
/product="KIAA1083 protein"
/protein_id="BAA83035.1"
/db_xref="GI:5689503"
/translation="MNSPGGRGKKKSGGASNPVPRPPPCLLPAPPAAGPAPPES
PHKRLNYFSYPLFVGFALLRLVAFHGLGLFWLQCFRFRALMAKRSAGAPAPASA
SAPAPVGGGEAERVVPHKQAFYISIALRIDEDKAGQKEQAVENYKKGIELEKGI
AVITGQGEQECERARRLOAKMTNLVMAKDRLOJLESQVAPVKKDPJTHISNLSRPSK
ACRNCHQSESGAVPKXKDPJTHISNLSRPSKTYKTSAGLSHRRAPSYSLSMVS
GVKQSGAPATTHKGFPTKTRNKPSPTTATRKKKDLKRNVDSEDLNINVEIVD
NGTAVRDEDDIAGQDLAKQALQEIIVILPSLRPELFGRLAPARGLLFGPPGNGKMLA
KAAVESNATFFNISASLTSKYVGESEKLVRAFAVARELQSEIIFIDEVDSLLCER
REGEDASRLKTEPLIEFDGVSAGDERVLVAGATNRPQSEDAVLRFPIKRVVYSL
PNEETRLLLKLLCKQGSPLTQKEAQLARMTDQYSGSDTLAKIAALGPIRELFK
ECVKNYSASEMRNIRLSDFTESLKKIKRVSYPOTLEAVIRWNKDFGCTTV"
BASE COUNT 1532 a 943 c 1077 g 1568 t
ORIGIN

Query Match 94.2%; Score 472; DB 9; Length 5120;
Best Local Similarity 100.0%; Pred. No. 1.7e-69;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGTTCGGTGGTCTGCGGAGGCGGGTTATGGCGGCGGCGGAGTGAGAGTGTGAATG 60
|||
Db 165 GGTTCGGTGGTCTGCGGAGGCGGGTTATGGCGGCGGCGGAGTGAGAGTGTGAATG 224
|||
QY 61 AATTCTCCGGTGGAGGAGGAGAAAGGCTCGCGGCGGCGGAGCAACCGGTGCCT 120
|||
Db 225 AATTCTCCGGTGGAGGAGGAGAAAGGCTCGCGGCGGCGGAGCAACCGGTGCCT 284
|||


```
QY 480 GGGGGCTGGGGAGGGGGGGC 501
      |||||
Db 23262 GGGGGCTGGGGAGGGGGGGC 23283

RESULT 9
AX093576 2152 bp DNA linear PAT 30-MAR-2001
LOCUS Sequence 136 from Patent WO0118198.
ACCESSION AX093576
VERSION AX093576.1 GI:13510014
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1
AUTHORS Weissbach, J. and Hagan, J.
TITLE Cloning, expression and characterization of the spg4 gene
responsible for the most frequent form of autosomal spastic
paraplegia
JOURNAL Patent: WO 0118198-A 106 15-MAR-2001;
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)
FEATURES
    source
        location/Qualifiers
            1..2152
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /note="APN Spg4 complete"
BASE COUNT 561 a 552 c 563 g 456 t
ORIGIN
    Query Match 69.6%; Score 342.8; DB 6; Length 2152;
    Best Local Similarity 86.7%; Pred. No. 7.3e-49;
    Matches 409; Conservative 0; Mismatches 57; Indels 6; Gaps 2;

QY 1 GGTTCCTCCGTCTGCGGAGGCGGGTATGGGGGGGGGCGAGTGAGAGCTGTGAATG 60
      |||||
Db 79 GGTTCCTCCGTCTGCGGAGGCGGGTATGGGGGGGGGCGAGTGAGAGCTGTGAATG 138

QY 61 AATTCTCCGGTGGACGAGGGAAGAAAGGCTCGGGAGGSCCAGCAACCGGTGCT 120
      |||||
Db 139 AGTTCTCCGGCGGACGAGGGAAGAAAGGCTCGGGAGGSCCAGCAACCGGTGCT 195

QY 121 CCCAGGCTCCGCCCCCTTGCTGGCCCCCGCCCCCTCCCGCGCGCGCGCCCCCTCG 180
      |||||
Db 196 GCGAGGCTCCGCCCCCGCGGGTCCCGCCCCCT---GCGCGCGCGCGCCCCCTGCG 252

QY 181 CCGGAGTCGCGGCATAGCGGAACCTGTACTATTCTCTACCCGCTGTTGTAGGCTTC 240
      |||||
Db 253 GCGGGCTCGCGGCTTAAGCGGAACCTGTCTTTCTCTCCCGCTGCTCGGCTTC 312

QY 241 GCGCTGCTGGTTTGGTGGCTTCACCTGGGGCTCTCTCTGTTGTGGCTCTGCCAGCGC 300
      |||||
Db 313 GCGCTGCTGGCTGCTGGCTTCCACCTGGGGCTCTCTCTGTTGTGGCTCTGCCAGCGC 372

QY 301 TTCTCCGCGCCCTCATGGGAGCCCAAGAGAGCTCCGGGGCGGGCGGCGACCTGCTCG 360
      |||||
Db 373 TTCTCCGCGCCCTCATGGGAGCCCAAGAGAGCTCCGGAGCCCGGCGGCGCGCGCTCG 432

QY 361 GCTTCGGCCCCCGCGCGGTGCGGGGGGGAGGCGGAGGCGCTCCGAGCTCTTCCAGAA 420
      |||||
Db 433 GCTTCGGCCCCCGCGGAGGCTGGGGAGGCGGAGGCTCCGAGGCTCCGAGCTCTTCCAG 492

QY 421 CAGGCTTCGAGTACATCTCCATTGCCCTGCGCATCGATGAGGATGAGAAAG 472
      |||||
Db 493 CAGGCTTCGAGTACATCTCCATTGCCCTGCGCATCGAGGAGGAGAAAG 544

RESULT 10
HSA325510 969 bp DNA linear PRI 16-JUL-2002
LOCUS Homo sapiens genomic sequence surrounding Not1 site, clone
DEFINITION HSA325510
          2116 bp mRNA linear ROD 10-FEB-2003
          YUS musculus, spastic paraplegia 4 homolog (human), clone MGC154785
          IMAGE:6441742, mRNA, complete cds.
ACCESSION BCC46286
VERSION BCC46286.1 GI:28279481

ACCESSION NB6-679R.
VERSION AJ325510
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 969)
AUTHORS Kutsenko, A.S., Gizatullin, R.Z., Al-Amin, A.N., Wang, F., Kvasha, S.M.,
Podowski, R.M., Matushkin, Y.G., Gyanchandani, A., Muravenko, O.V.,
Devitsky, V.G., Kochanov, N.A., Protopopov, A.I., Kashiba, V.I.,
Kisselev, L.L., Wasserman, W., Wahlestedt, C. and Zabarovsky, E.R.
TITLE Not1 flanking sequences: a tool for gene discovery and verification
of the human genome
JOURNAL Nucleic Acids Res. 30 (14), 3163-3170 (2002)
KEYWORDS
PUBMED 12136098
REFERENCE 2 (bases 1 to 969)
AUTHORS Zabarovsky, E.R.
TITLE Direct Submission
JOURNAL Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
Karolinska Institute, Theorell's vag, 3, Box 280, Stockholm 171 77,
Sweden
FEATURES
    source
        location/Qualifiers
            1..969
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /clone="NB6-679R"
BASE COUNT 159 a 281 c 324 g 205 t
ORIGIN
    Query Match 69.5%; Score 348.2; DB 9; Length 969;
    Best Local Similarity 92.8%; Pred. No. 1.1e-48;
    Matches 316; Conservative 3; Mismatches 28; Indels 1; Gaps 1;

QY 1 GGTTCCTCCGTCTGCGGAGGCGGGTATGGGGGGGGGCGAGTGAGAGCTGTGAATG 60
      |||||
Db 220 GGTTCCTCCGTCTGCGGAGGCGGGTATGGGGGGGGGCGAGTGAGAGCTGTGAATG 279

QY 61 AATTCTCCGGTGGACGAGGGAAGAAAGGCTCGGGAGGSCCAGCAACCGGTGCT 120
      |||||
Db 280 AATTCTCCGGTGGACGAGGGAAGAAAGGCTCGGGAGGSCCAGCAACCGGTGCT 339

QY 121 CCCAGGCTCCGCCCCCTTGCTGGCCCCCGCCCCCTCCCGCGCGCGCGCCCCCTCG 180
      |||||
Db 340 CCCAGGCTCCGCCCCCTTGCTGGCCCCCGCCCCCTCCCGCGCGCGCGCCCCCTCG 399

QY 181 CCGGAGTCGCGGCATAGCGGAACCTGTACTATTCTCTACCCGCTGTTGTAGGCTTC 240
      |||||
Db 400 CCGGAGTCGCTGGCATAGCGGAACCTGTACTATTCTCTACCCGCTGTTGTAGGCTTC 459

QY 241 GCGCTGCTGGTTTGGTGGCTTCACCTGGGGCTCTCTCTGTTGTGGCTCTGCCAGCGC 300
      |||||
Db 460 GCGCTGCTGGTTTGGTGGCTTCACCTGGGGCTCTCTCTGTTGTGGCTCTGCCAGCGC 519

QY 301 TTCTCCGCGCCCTCATGGGAGCCCAAGAGAGCTCCGGGGCGGGCGGCGACCTGCTCG 360
      |||||
Db 520 TTGTTCCGCGCCCTCATGGGAGCCCAAGAGAGCTCCGGAGCGGGCGGCGGCACTGCTTG 579

QY 361 GCTTCGGCCCCCGCGCGGTGCGGGGGGGAGGCGGAGGCGCGGTC 405
      |||||
Db 580 GCTTCGGGCTCGGCGGCTTCGGGCGGG-GAGGCGGAGCGGTC 623

RESULT 11
BCC46286 2116 bp mRNA linear
LOCUS YUS musculus, spastic paraplegia 4 homolog (human), clone MGC154785
DEFINITION IMAGE:6441742, mRNA, complete cds.
ACCESSION BCC46286
VERSION BCC46286.1 GI:28279481
```

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REMARK
COMMENT

MGC.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2116)
Straussberg, R.
Direct Submission
Submitted (31-JAN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2850,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgaps-asm@nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNLS)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.gen@nih.gov
Akhtar, N., Ayala, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Lalic, P., Legaspi, R.,
Maduro, Q.L., Masello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W.,
Tsurganov, C., Vogt, C.L., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.
Clon distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/ILNLS at: <http://image.llnl.gov>
Series: IRAX Plate: 100 Row: 1 Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, GenomeScan gene prediction, similarity but not identity
to protein.
Location/Qualifiers
1..2116
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="MGC:54756 IMAGE:6441742"
/tissue type="Mammary tumor, XYTV-175/INT3 model, 5 month
old mouse. Taken by biopsy."
/clone_lib="NCI CGAP_Mam2"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
Size: 1892
/codon_start=1
/product="spastic paraplegia 4 homolog (human)"
/protein_id="AAH46286.1"
/db_xref="GI:28273482"
/db_xref="LocusID:50850"
/translation="MSSPAGRRKKKSGGASPPAPRPPPPAVPAPAPAGPAPAGSPSP
KRPSPSPSPVVGFAFLRLJLACHLGLEFAWLCQPSRALMAKSGSGTAPAPASPS
PAPGGAESVRVFWQAFYISIAIDDEENKGEKQAVSWYKKGLEENKGIATY
VTGGQYQVARRLQAKNMNINLVNAKRLQLLEKLPVLOFSKSDVYNESNTLTP
NGHLQSEGAVPKPKDPLTHASNSLPSKTVLKSAGLSGHHRRAPSCGLSNVSGAR
PGFGPAATTHKGTPKNRTNKSTPTAVRKKKQKPNFVDSNANLTYNEVDNGT
AVKFDIAGQELAKQALQEIIVLPSPPELFTGLRAPARGLLFPFGNGKTMKARV
AAESNATFNISAALTSKYVGEKLVRALEAVARELOPSIIFIDEVESLLCEPREG
EHDASRLKTEFLIEFDGVSAGDDRVLVNGATNRCLEDAVLRPFIRRVVSLPNE
ETRLLLNLLCKQSPILQKELAQARMTDGYSGSLTALAKDALGPPIREKPEQV
KNMSASEMRNIRLSDFTELSKKIKRSVSPOTLEAYFWNKDFGDTIV"

BASE COUNT 610 a 515 c 546 g 445 t
ORIGIN

Query Match 69.1%; Score 346; DB 10; Length 2116;
Best local Similarity 86.1%; Pred. No. 2.1e-48;

Matches 408; Conservative 0; Mismatches 60; Indels 6; Gaps 2;
QY 8 GTGGTCTGGGAGGGGGTATGGGGGGGGGGAGTGGAGAGCTGTGATGATTC 67
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1 GTGGTCTGGGAGGGGGTATGGGGGGGGGGAGTGGAGAGCTGTGATGATTC 60
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 69 CGGTGGACGAGGAGAGAGAGGCTCGGGGGGGGGAGAGAGAGAGAGAGG 127
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 CGGGGAGAGAGAGAGAGAGAGGCTCGGGGGGGGGAGAGAGAGAGAGAG 117
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 128 CTGGGCGGGGGTGGGCTGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGG 187
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 116 CTGGGCGGGGGTGGGCTGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGG 174
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 186 CGGGGATAGAGAGAGAGAGAGGCTGTGATGAGAGAGAGAGAGAGAGAG 247
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 175 CGGGGCTAGAGAGAGAGAGAGAGGCTGTGATGAGAGAGAGAGAGAGAG 234
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 248 TGGGTTGGTGGGCTTCCAGCTGGGGGGGGGGGGGGGGGGGGGGGGGGGG 307
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 235 TGGGCTGGTGGGCTGGGCTGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGG 294
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 308 GCGGCTCATGGGAGAGAGAGAGAGGCTCGGGGGGGGGGGGGGGGGGGGG 367
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 295 GCGGCTCATGGGAGAGAGAGAGAGAGGCTCGGGGGGGGGGGGGGGGGGG 354
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 369 CCGGGGGGGGGTGGGCTGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 427
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 355 CCGGAGCGGGGGAGAGGGGTGGGAGGGGGGGGGGGGGGGGGGGGGGGGG 414
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 428 TGGAGTACATGCTGCTTGGGCTGGGCGAGAGAGAGAGAGAGAGAGAGAG 481
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 415 TGGAGTACATGCTGCTTGGGCTGGGCGAGAGAGAGAGAGAGAGAGAGAG 465
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
AC101733
LOCUS
DEFINITION Mus musculus clone RP23-417L6, LOW-PASS SEQUENCE SAMPLING.
AC101733
VERSION AC101733.1 GI:17060508
KEYWORDS HTG; HTGS PHASE0.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
1 (bases 1 to 71793)
REFERENCE Birren, B., Linton, L., Nussbaum, C. and Lander, E.
AUTHORS Birren, B., Linton, L., Nussbaum, C. and Lander, E.
TITLE Mus musculus, clone RP23-417L6
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 71793)
AUTHORS Birren, B., Linton, L., Nussbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., DeArctiano, K., Dewar, K., Diaz, C.S., Dodge, S., Faro, S.,
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Jarocque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, C., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, C.,
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.G., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (23-NOV-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02142, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L1731
Center clone name: 477_L_6

* NOTE: This record contains 91 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 590: contig of 693 bp in length
* 591: gap of 100 bp
* 791: contig of 673 bp in length
* 1466: gap of 100 bp
* 1566: gap of 100 bp
* 2271: contig of 705 bp in length
* 2372: gap of 100 bp
* 3066: contig of 695 bp in length
* 3166: gap of 100 bp
* 3851: contig of 685 bp in length
* 3951: gap of 100 bp
* 4636: contig of 685 bp in length
* 4737: gap of 100 bp
* 5437: gap of 100 bp
* 5537: contig of 698 bp in length
* 6235: gap of 100 bp
* 6335: contig of 693 bp in length
* 7034: gap of 100 bp
* 7134: contig of 688 bp in length
* 7822: gap of 100 bp
* 7922: contig of 683 bp in length
* 8611: gap of 100 bp
* 8711: contig of 701 bp in length
* 9412: gap of 100 bp
* 9512: contig of 719 bp in length
* 10231: gap of 100 bp
* 10331: contig of 711 bp in length
* 11042: gap of 100 bp
* 11442: contig of 705 bp in length
* 11847: gap of 100 bp
* 11946: contig of 695 bp in length
* 12641: contig of 689 bp in length
* 12741: gap of 100 bp
* 12742: contig of 689 bp in length
* 13431: gap of 100 bp
* 13531: contig of 687 bp in length
* 14217: gap of 100 bp
* 14218: contig of 691 bp in length
* 14318: contig of 670 bp in length
* 15009: gap of 100 bp
* 15109: contig of 670 bp in length
* 15779: gap of 100 bp
* 15879: contig of 693 bp in length
* 16571: gap of 100 bp
* 16572: contig of 713 bp in length
* 17384: gap of 100 bp
* 17485: contig of 681 bp in length
* 18165: gap of 100 bp
* 18265: contig of 693 bp in length
* 18958: contig of 693 bp in length
* 18959: gap of 100 bp
* 19058: contig of 693 bp in length
* 19059: gap of 100 bp
* 19757: contig of 693 bp in length
* 19857: gap of 100 bp
* 20510: contig of 653 bp in length
* 20610: gap of 100 bp
* 21295: contig of 685 bp in length
* 21395: gap of 100 bp
* 22087: contig of 692 bp in length
* 22187: gap of 100 bp
* 22860: contig of 673 bp in length
* 22960: gap of 100 bp
* 23646: contig of 686 bp in length
* 23746: gap of 100 bp
* 24429: contig of 683 bp in length
* 24529: gap of 100 bp
* 25212: contig of 683 bp in length
* 25312: gap of 100 bp
* 25998: contig of 686 bp in length
* 26098: gap of 100 bp
* 26800: contig of 702 bp in length
* 26900: gap of 100 bp
* 27619: contig of 719 bp in length
* 27719: gap of 100 bp
* 28436: contig of 717 bp in length
* 28536: gap of 100 bp
* 29237: contig of 701 bp in length
* 29337: gap of 100 bp
* 30032: contig of 695 bp in length
* 30132: gap of 100 bp
* 30819: contig of 687 bp in length
* 30919: gap of 100 bp
* 31590: contig of 671 bp in length
* 31690: gap of 100 bp
* 32367: contig of 677 bp in length
* 32467: gap of 100 bp
* 33142: contig of 675 bp in length
* 33242: gap of 100 bp
* 33908: contig of 666 bp in length
* 34008: gap of 100 bp
* 34697: contig of 689 bp in length
* 34797: gap of 100 bp
* 35465: contig of 668 bp in length
* 35565: gap of 100 bp
* 36268: contig of 703 bp in length
* 36368: gap of 100 bp
* 37073: contig of 705 bp in length
* 37173: gap of 100 bp
* 37858: contig of 685 bp in length
* 37958: gap of 100 bp
* 38636: contig of 680 bp in length
* 38738: gap of 100 bp
* 39436: contig of 698 bp in length
* 39536: gap of 100 bp
* 40211: contig of 675 bp in length
* 40311: gap of 100 bp
* 41010: contig of 699 bp in length
* 41110: gap of 100 bp
* 41811: contig of 701 bp in length
* 41911: gap of 100 bp
* 42612: contig of 701 bp in length
* 42712: gap of 100 bp
* 43407: contig of 695 bp in length
* 43507: gap of 100 bp
* 44181: contig of 674 bp in length
* 44281: gap of 100 bp
* 44964: contig of 683 bp in length
* 45064: gap of 100 bp
* 45778: contig of 714 bp in length
* 45878: gap of 100 bp
* 46577: contig of 699 bp in length
* 46678: gap of 100 bp
* 47362: contig of 685 bp in length
* 47462: gap of 100 bp

```
* 47463 48154: contig of 692 bp in length
* 48155 48254: gap of 100 bp
* 48255 48953: contig of 699 bp in length
* 48954 49053: gap of 100 bp
* 49054 49736: contig of 685 bp in length
* 49739 49838: gap of 100 bp
* 49839 50526: contig of 688 bp in length
* 50527 50626: gap of 100 bp
* 50627 51321: contig of 695 bp in length
* 51322 51421: gap of 100 bp
* 51422 52113: contig of 692 bp in length
* 52114 52213: gap of 100 bp
* 52214 52909: contig of 696 bp in length
* 52910 53009: gap of 100 bp
* 53010 53659: contig of 662 bp in length
* 53660 53769: gap of 100 bp
* 53770 54462: contig of 693 bp in length
* 54463 54562: gap of 100 bp

Query Match      54.7%  Score 274;  DB 2;  Length 71293;
Best Local Similarity 81.8%  Pred. No. 1.1e-36;
Matches 342;  Conservative 0;  Mismatches 70;  Indels 6;  Gaps 2;

QY 1 GGTTCCTCGGTCGTGCGGAGAGCGGGTTATGCGGGCGGGCGGAGTGAGAGCTGTGAATG 60
   |||||
Db 57132 GGTTCCTCGGTCGTGCGGAGAGCGGGTTATGCGGGCGGGCGGAGTGAGAGCTGTGAATG 57191

QY 61 AATTCTCCGGTGCAGGAGGAGAGAAAGGTCGCGGGCGGCGGCGGAGCAACCGGTGCCT 120
   |||||
Db 57192 AGTTCTCCGGCCGACGACGAGCGAAGAAAGGTCGCGGGCGGCGGAG---CGCGGCGCC 57248

QY 121 CCGAGGCTCCGCGCCCTTGCTGCGCCCGCCCGCCCTCCCGCGGCGGCGGCGGCGGCGGCTCCG 180
   |||||
Db 57249 GCGAGGCTCCGCGCCCGCCCGCGGCGGCTCCCGCGGCGGCGGCGGCGGCGGCGGCTCCG 57305

QY 181 CCGAGTCCCGCATAGCGGAACTGTACTATTCCTCTACCGGCTGTTGTAGGCTTC 240
   |||||
Db 57306 GCGGCTCCGCGCTATGCGGCAAGAGAGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 57365

QY 241 GCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
   |||||
Db 57366 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 57425

QY 301 TTCTCCCGCGCCCTCATGCGGAGAGAGAGAGTCTGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCTCG 360
   |||||
Db 57426 TTCTCCCGCGCCCTCATGCGGAGAGAGAGAGTCTGCGGGAGAGAGAGTCTGCGGGCGGCGGCTCG 57485

QY 361 GCGTCCGCGCGCGCGCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTCCACA 419
   |||||
Db 57486 CCGTCCGCGCGCGAGAGCGGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTCCACA 57547

RESULT 13
HSA342191
LOCUS
DEFINITION
Homo sapiens genomic sequence surrounding NotI site, clone
HSJ-6AN5RS.
ACCESSION
AJ342191
VERSION
AJ342191.1 GI:15886603
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1. (bases 1 to 721)
AUTHORS
Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F., Kvasha,S.M.,
Podowski,R.M., Matushkin,Y.G., Gyanchandani,A., Muravenko,O.V.,
Levitsky,V.G., Kolchanov,N.A., Protodopov,A.I., Kashuba,V.I.,
Kisselev,L.L., Wasserman,W., Wahlestedt,C. and Zabarovskiy,E.R.
NotI flanking sequences: a tool for gene discovery and verification
of the human genome
Nucleic Acids Res. 30 (14), 3163-3170 (2002)
JOURNAL
MEDLINE
22131767
PUBMED
12136098
REFERENCE
2 (bases 1 to 712)
AUTHORS
Zabarovskiy,E.R.
Direct Submission
Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
Sweden
FEATURES
Location/Qualifiers
1..712
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="HSJ-6AN5RS"
BASE COUNT 115 a 215 c 260 g 128 t 3 others
ORIGIN
```

```
PUBMED 12136098
REFERENCE 2 (bases 1 to 721)
AUTHORS
Zabarovskiy,E.R.
Direct Submission
Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
Sweden
FEATURES
Location/Qualifiers
1..721
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="HSJ-6AN5RS"
BASE COUNT 115 a 215 c 260 g 128 t 3 others
ORIGIN

Query Match      49.4%  Score 247.4;  DB 9;  Length 721;
Best Local Similarity 95.0%  Pred. No. 7.4e-32;
Matches 266;  Conservative 0;  Mismatches 12;  Indels 2;  Gaps 1;

QY 1 GGTTCCTCGGTCGTGCGGAGAGCGGGTTATGCGGGCGGGCGGAGTGAGAGCTGTGAATG 60
   |||||
Db 220 GGTTCCTCGGTCGTGCGGAGAGCGGGTTATGCGGGCGGGCGGAGTGAGAGCTGTGAATG 279

QY 61 AATTCTCCGGTGCAGGAGGAGAGAAAGGTCGCGGGCGGCGGCGGAGCAACCGGTGCCT 120
   |||||
Db 280 AATTCTCCGGTGCAGGAGGAGAGAAAGGTCGCGGGCGGCGGCGGAGCAACCGGTGCCT 339

QY 121 CCGAGGCTCCGCGCCCTTGCTGCGCCCGCCCGCCCTCCCGCGGCGGCGGCGGCGGCGGCTCCG 180
   |||||
Db 340 TCCAGGCTCCGCGCCCTTGCTGCGCCCGCCCGCCCTCCCGCGGCGGCGGCGGCGGCTCCG 399

QY 181 CCGAGTCCCGCATAGCGGAACTGTACTATTCCTCTACCGGCTGTTGTAGGCTTC 240
   |||||
Db 490 CCGAGTCCCGCATAGCGGAACTGTACTATTCCTCTACCGGCTGTTGTAGGCTTC 459

QY 241 GCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 283
   |||||
Db 460 GCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 497

RESULT 14
HSA342199
LOCUS
DEFINITION
Homo sapiens genomic sequence surrounding NotI site, clone
HSJ-6AN4RS.
ACCESSION
AC342199
VERSION
AC342199.1 GI:15886616
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1. (bases 1 to 712)
AUTHORS
Kutsenko,A.S., Gizatullin,R.Z., Al-Amin,A.N., Wang,F., Kvasha,S.M.,
Podowski,R.M., Matushkin,Y.G., Gyanchandani,A., Muravenko,O.V.,
Levitsky,V.G., Kolchanov,N.A., Protodopov,A.I., Kashuba,V.I.,
Kisselev,L.L., Wasserman,W., Wahlestedt,C. and Zabarovskiy,E.R.
NotI flanking sequences: a tool for gene discovery and verification
of the human genome
Nucleic Acids Res. 30 (14), 3163-3170 (2002)
JOURNAL
MEDLINE
22131767
PUBMED
12136098
REFERENCE
2 (bases 1 to 712)
AUTHORS
Zabarovskiy,E.R.
Direct Submission
Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
Sweden
FEATURES
Location/Qualifiers
1..712
/organism="Homo sapiens"
/mol_type="genomic DNA"
JOURNAL
MEDLINE
22131767
PUBMED
12136098
REFERENCE
2 (bases 1 to 712)
AUTHORS
Zabarovskiy,E.R.
Direct Submission
Submitted (16-MAY-2001) Microbiology and Tumorbiology Centre,
Karolinska Institute, Theorells vag, 3, Box 280, Stockholm 171 77,
Sweden
```



```
FT exon 35749..35832
FT /*tag= f
FT /number= "3"
FT intron 35833..45022
FT /*tag= g
FT /number= "3"
FT exon 45023..45118
FT /*tag= h
FT /number= "4"
FT intron 45119..46063
FT /*tag= i
FT /number= "4"
FT exon 46064..46051
FT /*tag= j
FT /number= "5"
FT intron 46052..461927
FT /*tag= k
FT exon 461928..462061
FT /number= "5"
FT intron 462062..462344
FT /*tag= m
FT exon 462345..462438
FT /*tag= n
FT intron 462439..463173
FT /*tag= o
FT exon 463174..463248
FT /*tag= p
FT intron 463249..464533
FT /*tag= q
FT exon 464534..464705
FT /*tag= r
FT intron 464706..462788
FT /*tag= s
FT exon 462789..462864
FT /*tag= t
FT intron 462865..463122
FT /*tag= u
FT exon 463123..463194
FT /*tag= v
FT intron 463195..463334
FT /*tag= w
FT exon 463335..463414
FT /*tag= x
FT intron 463415..463429
FT /*tag= y
FT exon 463430..463472
FT /*tag= z
FT intron 463473..463561
FT /*tag= aa
FT exon 463562..463641
FT /*tag= ab
FT intron 463642..463662
FT /*tag= ac
FT exon 463663..463733
```

```
FT /*tag= ad
FT /number= "15"
FT intron 91234..93443
FT /*tag= ae
FT exon 93444..93484
FT /*tag= af
FT intron 93485..100593
FT /*tag= ag
FT exon 100600..102009
FT /*tag= ah
FT /number= "17"
XX FR2798138-A1.
XX 09-MAR-2001.
XX 03-SEP-1999; 99FR-0311097.
XX 03-SEP-1999; 99FR-0311097.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX Weissenbach J, Hazan J;
XX WPI; 2001-283966/30.
XX
XX New human nucleic acid from the SPG4 gene, useful e.g. for diagnosis of
XX autosomal dominant familial spastic paraplegia and in drug screening -
XX
XX Claim 2; Page 45-106; 145pp; French.
XX
XX The present sequence represents a human SPG4 gene. The SPG4 gene encodes
XX a spastin polypeptide. Mutations in the SPG4 gene are responsible for
XX autosomal dominant familial spastic paraplegia. SPG4 polynucleotides,
XX and their fragments, are used to screen DNA banks for sequences that
XX encode spastin (particularly sequences in other mammals, specifically
XX mice); to identify SPG4 mutations, or other genetic anomalies.
XX particularly for diagnosis of autosomal dominant familial spastic
XX paraplegia (PSF-AD); to identify promoters and other regulatory elements
XX of the SPG4 gene; for detection and amplification; for recombinant
XX production of spastin; and for diagnostic genotyping of PSF-AD.
XX
XX Sequence 110000 BP; 35622 A; 21640 C; 22817 G; 14921 T; 0 other;
XX
XX Query Match 100.0%; Score 501; DB 22; Length 110000;
XX Best Local Similarity 100.0%; Pred. NO. 8.7e-92;
XX Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 GGTTCCGTCGGTCTGCGGGAGGCGGGTTATGCGGGCGGCGGAGTGAGAGCTGTGAATG 60
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 10000 GGTTCCGTCGGTCTGCGGGAGGCGGGTTATGCGGGCGGCGGAGTGAGAGCTGTGAATG 10059
XX
XX 61 AATTCTCCGGGTGGACGAGGGAGAGAAAGGTCGGGCGGCGGCGGAGCAAGCGTGCCT 120
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 10060 AATTCTCCGGGTGGACGAGGGAGAGAAAGGTCGGGCGGCGGCGGAGCAAGCGTGCCT 10119
XX
XX 121 CCCAGGCGCTCCGCGCCCTTGCCTGGGCGCCCGCCCTCCGCGGCGGCGGCGGCGCTCCG 180
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 10120 CCCAGGCGCTCCGCGCCCTTGCCTGGGCGCCCGCCCTCCGCGGCGGCGGCGGCGCTCCG 10179
XX
XX 181 CCCAGATCGCCGGATAAGCGGAACCTGTACTATTTCCTCTACCGGCTGTTTGTAGGCTTC 240
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 10180 CCCAGATCGCCGGATAAGCGGAACCTGTACTATTTCCTCTACCGGCTGTTTGTAGGCTTC 10239
XX
XX 241 GCGCTGCTCGGTTTGGTCGCTTCCACCTGGGAGCTCCTCTTCGTGTGGCTTGGCAGCGC 300
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
XX 10240 GCGCTGCTCGGTTTGGTCGCTTCCACCTGGGAGCTCCTCTTCGTGTGGCTTGGCAGCGC 10299
XX
XX 301 TTCTCCGCGCCCTCATGGCAGCAAGAGAGCTCCGGGCGGCGGCGGAGCAAGCTGCCTCG 360
XX ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

Db 10300 TTCTCCGCGCCCTCATGTGGAGGCAAGAGGAGCTCCGGGGCGGCGCCAGACCTGCCTCG 10359
QY 361 GCCTCGGCCCCCGGCGCGGTGCGGGGGGAGCGCGAGCGCGTCCGAGCTTCCACAAA 420
Db 10360 GCCTCGGCCCCCGGCGCGGTGCGGGGGGAGCGCGAGCGCGTCCGAGCTTCCACAAA 10419
QY 421 CAGGCTTCGAGTACATCTCCATTCGCCCTCGGCATCGATGAGGATGAGAAAGGTAAGTAG 480
Db 10420 CAGGCTTCGAGTACATCTCCATTCGCCCTCGGCATCGATGAGGATGAGAAAGGTAAGTAG 10479
QY 481 GGGGCTGGGGGAGGGGGCGGC 501
Db 10480 GGGGCTGGGGGAGGGGGCGGC 10500

RESULT 2
AAF848C1
ID AAF848C1 standard; cDNA; 3263 BP.
XX
AC AAF848C1;
DT 09-JUL-2001 (first entry)
XX
DE Nucleotide sequence of the human SPG4 polypeptide.
XX
KW Human; SPG4 gene; spastin; PSF-AD; gene therapy;
KW autosomal dominant familial spastic paraplegia; as
XX
CS Homo sapiens.
XX
FH Key
FT CDS
FT Location/Qualifiers
FT 126..1976
FT /*tag= a
FT /product= "spastin"
XX
PN FR2799138-A1.
XX
PD 09-MAR-2001.
XX
PF 03-SEP-1999; 99FR-0011097.
XX
PR 03-SEP-1999; 99FR-0011097.
XX
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Weissenbach J; Hazan J;
XX
PR WP1; 2001-283966/30.
XX
PR P-PSDB; AAB68137.
XX
PT New human nucleic acid from the SPG4 gene, useful e.g. for diagnosis of
XX autosomal dominant familial spastic paraplegia and in drug screening -
XX
PS Claim 2; Page 106-108; 145pp; French.
XX
CC The present sequence encodes a human spastin polypeptide. Spastin is
XX encoded by the SPG4 gene. Mutations in the SPG4 gene are responsible for
XX autosomal dominant familial spastic paraplegia. SPG4 polynucleotides,
XX and their fragments, are used to screen DNA banks for sequences that
XX encode spastin (particularly sequences in other mammals, specifically
XX mice); to identify SPG4 mutations, or other genetic anomalies,
XX particularly for diagnosis of autosomal dominant familial spastic
XX paraplegia (PSF-AD); to identify promoters and other regulatory elements
XX of the SPG4 gene; for detection and amplification; for recombinant
XX production of spastin; and for diagnostic genotyping of PSF-AD.
XX
SQ Sequence 3263 BP; 956 A; 664 C; 703 G; 940 T; 0 other.

Query Match 94.23; Score 472; DB 22; Length 3263;
Best Local Similarity 100.0%; Pred. No. 5.2e-86;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGTTCCTCGCTCGGGGAGCGGGGTTATGGCGGCGGCGGAGTGGAGGCTGTGAATG 50

```

```

Db 59 GGTTCCTCGCTCGGGGAGCGGGTTATGGCGGCGGCGGAGTGGAGGCTGTGAATG 128
QY 61 AATTCTCCGGTGGAGGAGGAAGAAAGGCTCCGGGGCGGCGGAGCAACCCGTCCT 120
Db 129 AATTCTCCGGTGGAGGAGGAAGAAAGGCTCCGGGGCGGCGGAGCAACCCGTCCT 188
QY 121 CCCAGGCTCCGCCCTTGCTTGGCCCGCCCGCCCTCCCGCGGCGGCGGCGGCGGCGG 180
Db 189 CCCAGGCTCCGCCCTTGCTTGGCCCGCCCGCCCTCCCGCGGCGGCGGCGGCGGCGG 248
QY 191 CCCAGTCCCGCATAGCGGAACCTTACTATTCTCTACCCGCTGTTTGTAGCTTC 240
Db 249 CCCAGTCCCGCATAGCGGAACCTTACTATTCTCTACCCGCTGTTTGTAGCTTC 308
QY 241 GCGCTGCTCGCTTGGTCCCTTCCACTGGGGCTCTCTTGGTGGCTCTCCAGCGC 300
Db 309 GCGCTGCTCGCTTGGTCCCTTCCACTGGGGCTCTCTTGGTGGCTCTCCAGCGC 368
QY 391 TTCTCCCGGCCCTCATGCGAGCCAAAGAGAGCTCCGGGGCGGCGGAGCTTCCAGCTCG 360
Db 369 TTCTCCCGGCCCTCATGCGAGCCAAAGAGAGCTCCGGGGCGGCGGAGCACTCGCTCG 428
QY 361 CCTCGGCGCGGCGCGGTGCGGGGGGAGGCGGCGGCGGCGGCTCCAGTCTTCCACAAA 420
Db 429 CCTCGGCGCGGCGCGGTGCGGGGGGAGGCGGCGGCGGCGGCTCCAGTCTTCCACAAA 488
QY 421 CAGGCTTCGAGTACATCTCCATTCGCCCTCGGCATCGATGAGGATGAGAAAG 472
Db 489 CAGGCTTCGAGTACATCTCCATTCGCCCTCGGCATCGATGAGGATGAGAAAG 540

RESULT 3
ABQ39732/C.
ID ABQ39732 standard; DNA; 777 BP.
XX
AC ABQ39732;
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 26323.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
CS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
XX
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIC-) EPICENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
PR WP1; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
XX for diagnosis and prognosis, comprises selective hybridization of
XX amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert

```



```
XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX CC This invention describes a novel method for determining the degree of
XX CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX CC genomic sample of DNA. The sample is treated chemically to convert
XX CC cytosine (C) but not methylated C, to uracil, then part of the genomic
XX CC DNA that contains the target C is amplified to form a labeled amplicon.
XX CC The amplicon is hybridised to two classes, each with at least one
XX CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
XX CC and the degree of hybridisation to both classes is determined from the
XX CC label on the amplicon. From the ratio of labels hybridised to the two
XX CC classes of oligomers, the degree of methylation is calculated. The method
XX CC is used: (i) for diagnosis and/or prognosis of side effects of
XX CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
XX CC of the central nervous, cardiovascular, gastrointestinal and respiratory
XX CC systems etc., particularly by detecting mutations or single nucleotide
XX CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
XX CC types and for investigating cell differentiation. The method allows the
XX CC methylation status of many C residues to be determined simultaneously.
XX CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
XX CC method for determining the degree of cytosine methylation described in
XX CC the disclosure of the invention.
XX SQ Sequence 777 BP; 317 A; 275 C; 79 G; 106 T; 0 other;

Query Match: 60.4%; Score 302.6; DB 24; Length 777;
Best Local Similarity 75.2%; Pred. No. 5.3e-52;
Matches 377; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 1 GGTTCCTCGTCTGCGGAGGCGGTATATGCGGGCGGGCGGAGTGAGAGCTGTGATG 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 AATTCTCCGCTGGACGAGGGAAGAAGAAAGCTCCGGCGCGCCAGCAACCCGGTGCCT 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 610 AATTCTCCGCTGGACGAGGGAAGAAGAAAGCTTCGGCGCGGTAGTAATTCGGTGT 551
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 CCCAGGCTCCGCCGCCCTTGTGCTGGCCCGCCCTCCGCCCGCGGGCGCCCTCG 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 550 TTTAGGTTTTCGTTTTTGTGTTTTCGTTTTTTCGTTTTTTCGTTTTTTCG 491
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 CCCAGTCCGCGCATAGCGGAACCTGTACTATTCTCTACCGGCTGTTTTCAGGCTTC 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 490 TTCAGTCTGCTGATAGCGGAATTTGTATTATTTTATTATTCGTTTTCAGGCTTC 431
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 GCGCTGCTGCTTGGTTCGCTTCCACCTGCGGCTCTCTCTCTGTTGCTGCGAGGCG 300
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 430 GCGTGTGCTTGGTTCGTTTTTATTTCGGGTTTTTTCGTTGCTGTTTTCAGGCT 371
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 TTCTCCCGGCGCTCATGCGAGCCCAAGAGGAGCTCCGGGCGCGCCGACACCTGCTCG 360
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 370 TTTTTCGCGTTTTTATGTTAGTTAAGAGGAGTTTCGGGCTCGGTTAGTATTGTTTCG 311
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 GCGTCCGCGCGCGCGCTGCTGCGGCGGCGAGCGAGCGCTCCGATCTTCCACAAA 420
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 310 GCTTCGCTTCGCGCTCGGTGCTCGGCGCGCGAGCTCGAGCGCTTCTTATATAA 251
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 CAGGCTTCAGTACATCTCATTTGCCCTGCGATCTCATGAGGATGAGAAAGGTACTAG 480
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 250 TAGGTTTCGAGTATATTTTATTTGTTTGGTATCTCATGAGGATGAGAAAGGTACTAG 191
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 461 GGGGCTGGGAGGGGGCGGC 501
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 190 GGGGTTGGGAGGGGGCGGC 170
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
ABQ24124/c
ID ABQ24124 standard; DNA; 562 BP.
XX AC ABQ24124;
XX XX
```

```
DT 12-JUL-2002 (first entry)
XX CC Oligonucleotide for detecting cytosine methylation SEQ ID NO 10715.
XX CC Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX CC drug; side effect; cancer; central nervous system; cardiovascular;
XX CC gastrointestinal; respiratory system; single nucleotide polymorphisms;
XX CC SNP; cell differentiation; ds.
XX OS Homo sapiens.
XX XX WO200218632-A2.
XX PN 07-MAR-2002.
XX PD 01-SEP-2001; 2001WO-EP13074.
XX PF 01-SEP-2000; 2000DE-1043826.
XX PR 05-SEP-2000; 2000DE-1044543.
XX XX (EPIG-) EPIGENOMICS AG.
XX PI Oleg A. Piepenhock C, Berlin K, Guelig D;
XX XX KPI; 2002-371829/40.
XX DR
XX CC Determining the degree of cytosine methylation in genomic DNA, useful
XX CC for diagnosis and prognosis, comprises selective hybridization of
XX CC amplicons from chemically treated DNA.
XX PS Claim 12; 56pp - Sequence Listing; 56pp; German.
XX CC This invention describes a novel method for determining the degree of
XX CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX CC genomic sample of DNA. The sample is treated chemically to convert
XX CC cytosine (C) but not methylated C, to uracil, then part of the genomic
XX CC DNA that contains the target C is amplified to form a labeled amplicon.
XX CC The amplicon is hybridised to two classes, each with at least one
XX CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
XX CC and the degree of hybridisation to both classes is determined from the
XX CC label on the amplicon. From the ratio of labels hybridised to the two
XX CC classes of oligomers, the degree of methylation is calculated. The method
XX CC is used: (i) for diagnosis and/or prognosis of side effects of
XX CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
XX CC of the central nervous, cardiovascular, gastrointestinal and respiratory
XX CC systems etc., particularly by detecting mutations or single nucleotide
XX CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
XX CC types and for investigating cell differentiation. The method allows the
XX CC methylation status of many C residues to be determined simultaneously.
XX CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
XX CC method for determining the degree of cytosine methylation described in
XX CC the disclosure of the invention.
XX SQ Sequence 562 BP; 113 A; 53 C; 195 G; 201 T; 0 other;

Query Match: 45.9%; Score 229.8; DB 24; Length 562;
Best Local Similarity 80.3%; Pred. No. 2.2e-37;
Matches 282; Conservative 0; Mismatches 67; Indels 2; Gaps 1;

QY 129 TCGGCCCCCTTCCTGGCCCCCGCCCCCTCCGCGCGCGCGCGCCCCCTCGGCGAGTC 188
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 562 TCGGCCCCCTTACCTAACCGCGCCCCCTCCGCGCGCGCGAGACCCCTCGGCGAATC 503
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 189 GCGGCATAGCGGAACCTGTACTATTTCCTCTACCGCTGTTTGTAGGCTTCGCGCTGCT 248
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 502 GCGGCATAAACGAACCTGTACTATTTCCTCTACCGCTATTATATAAATTCGCGCTACT 443
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 249 GCGTTTGGTTCGCTTCCACCTTCGCGGCTTCCTCTTCGTGTGGTGTCTGCCAGGCTTCTCCCG 308
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 442 AGCTTTAACTCGGCTTCACCTTAAACTCCTCTTCGTATTAAGTCTACCAAGGCTTCTCCCG 383
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 309 CGCCTCTAATGGCAGCAAGAGAGAGCTCCGCGCGCGCGCGACACCTGCTCGGCTCGGC 368
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


DB 382 CGCCCTCATACAAACCAAAATACTCCGAAACCCGCGCCACACCTACCTCGACTCGAC 323
QY 369 CCGCGCGCGGTGCGCGGGCGGAGCGCGGTCGAGTCTTCCACAAAGCGGCTT 428
DB 322 CC--GACGCGGATACCGGACGAAACGGAACGGTCCGATCTTCCACAAACACCTT 265
QY 429 CGAGTACATCTCCATTCGCTGCGCATGATGAGGATGAGAAAGGTAACCTA 479
DB 264 CGAATACATCTCCATTCGCTGCGCATGATGAGGATGAGAAAGGTAACCTA 214

RESULT 8
ID ABQ24125
XX
AC ABQ24125:
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 10712.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WC200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2000; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 562 BP; 20: A; 195 C; 53 G; 113 T; 0 other;
Query Match 45.9%; Score 229.8; DB 24; Length 562;

Best Local Similarity 80.3%; Pied. No. 2.12e-37;
Matches 282; Conservative C; Mismatches 67; Indels 2; Gaps 1;

QY 129 TCCGCGCCCTTGCTGGCCCGCCCTCCCGCGCGCGCGCGCGCGCGCGCGCGGTC 188
DB 1 TCCGCGCCCTTACCTAACCCCGCCCTCCCGCGCGCGCGCGCGCGCGCGCGCGGATC 63
QY 189 GCGGATAAGCGAACCTGTACTATTTCTCTACCCGCTGTTTSTAGGTTTCGGCTGCT 248
DB 61 GCGGATAAGCGAACCTGTACTATTTCTCTACCCGCTGTTTSTAGGTTTCGGCTGCT 120
QY 249 GCGTTTGGTTCGCTTCACCTGCGGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 308
DB 121 ACGTTTAAATCGCTTCACCTTAAACTCTCTCTGATAACTTACCTACCAACGCTTCTCCCG 180
QY 309 GCGCTTCATGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 368
DB 181 GCGCTTCATGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 240
QY 369 GCGGCGCGCGGTGCGCGGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 428
DB 241 CC--GACGCGGATACCGGACGAAACGGAACGGTCCGATCTTCCACAAACACCTT 298
QY 429 CGAGTACATCTCCATTCGCTGCGCATGATGAGGATGAGAAAGGTAACCTA 479
DB 299 CGAATACATCTCCATTCGCTGCGCATGATGAGGATGAGAAAGGTAACCTA 349

RESULT 9
ID ABQ24122
XX
AC ABQ24122:
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 10713.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WC200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
XX


```

/ FILING DATE: 27-FEB-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: KITS, Monica C
/ REGISTRATION NUMBER: 36,105
/ REFERENCE/DOCKET NUMBER: P1615-6007
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 202/638-5000
/ TELEFAX: 202/638-4910
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1269 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ ORIGINAL SOURCE:
/ ORGANISM: Streptomyces peuceletius
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1..1269
US-08-760-116-1

```

```

Query Match          9.9%; Score 49.4; DB 1; Length 1269;
Best Local Similarity 48.4%; Pred. No. 0.03;
Matches 137; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 108 CAACCCGGTGCCTCCAGGCTCCGCCCCCTGCTGCGCCGCGCCGCTCCGCGCGCG 167
DB 360 CGAGCGGACGGATCGCATCGCGGATCGCGCGCGCTGGCTGCTCAGCGAAGTCCGCGAGCC 419
QY 168 GCCGCGCCCTCCGCGCGAGTGGCGGCATAGCGGAACTGTACTATTTCCTCAGCGCT 227
DB 420 CTCGCGCGGTCGCGGCAACCGGCGAGCTGATGGCGGCTTGGCTAGCACTTCCGCT 479
QY 228 GTTTGTAGGCTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 287
DB 480 GTTGTGATCTGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 539
QY 288 GCTGTCCAGCGCTTCTCCGCGCGCTCATGGCGAGCAAGAGAGAGCTCCGAGGCGG 347
DB 540 CGTCAGCGTTCTCAGGCACTCGGCTCGGCGCGCTCGGCGCGCTCGGCGCGCTGAGCGGAC 599
QY 348 AGCACTGCTCCGCGCTCGGCGCGCTCGGCGCGCTCGGCGCGCTCGGCGCGCTCG 397
DB 600 GGACCTCGCGCGCGCTCGGCGAGACCTCGGCGCGCTCGGCGCGCTCGGCGAGAC 642

```

```

RESULT 4
US-08-102-863-10
/ Sequence 10, Application US/08102863
/ Patent No. 546659C
/ GENERAL INFORMATION:
/ APPLICANT: SARIASLANI, SIMA
/ TITLE OF INVENTION: CONSTITUTIVE
/ TITLE OF INVENTION: EXPRESSION OF P45CSOY
/ TITLE OF INVENTION: AND FERREDOXIN-SOY IN
/ TITLE OF INVENTION: STREPTOMYCES
/ NUMBER OF SEQUENCES: 12
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: E. I. DU PONT DE NEMOURS
/ ADDRESSEE: AND COMPANY
/ STREET: 1007 MARKET STREET
/ CITY: WILMINGTON
/ STATE: DELAWARE
/ COUNTRY: USA
/ ZIP: 19898
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0,

```

```

/ SOFTWARE: Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/102,863
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/07/807,001
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: GALLEGO, R. THOMAS
/ REGISTRATION NUMBER: 32,692
/ REFERENCE/DOCKET NUMBER: CR-9000
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 302-892-7342
/ TELEFAX: 302-892-7949
/ INFORMATION FOR SEQ ID NO: 10:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1735 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
US-08-102-863-10

```

```

Query Match          9.5%; Score 47.8; DB 1; Length 1735;
Best Local Similarity 47.3%; Pred. No. 0.068;
Matches 149; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

QY 127 CCTCCGCGCCCTTGCTGGCCCGCCGCGCCCTCCGCGCGCGCGCGCGCGCGCGAG 186
DB 205 CGGCGCGCGCGAGAACCTCGACCCACCTCCCGCGCGCGCGCGCGCGCGCGCGAG 264
QY 187 TGGCGCATTAAGCGGAACTGTACTATTCTCTACCGCGCTGTTTGTAGGGCTTCGCGCTG 246
DB 265 CGCGGTTGCGCTTACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 324
QY 247 CTGCGTTTGGTTCGCTTCCACCTGGGCGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTG 306
DB 325 AGCGGGTACCGCTTCTGACGCGACGCGCGCTCTGCGCGCGCTCACCGCGCGCGCGCGCG 384
QY 307 CGCGCCCTCATGGAGCAAGAGAGAGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCTG 366
DB 385 CGTGGCTATGGCGAGCGCGCGCGCTCTCACCGAGCGCGCGCGCGCGCGCGCGCGCTG 444
QY 347 GCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 426
DB 445 CGGCGCGAGCGGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 504
QY 427 TTCGAGTACATCTCC 441
DB 505 CCGAGCAACAACAC 519

```

```

RESULT 5
PCT-US92-10885-10
/ Sequence 10, Application PC/TUS9210885
/ GENERAL INFORMATION:
/ APPLICANT: SARIASLANI, SIMA
/ TITLE OF INVENTION: CONSTITUTIVE
/ TITLE OF INVENTION: EXPRESSION OF P45CSOY
/ TITLE OF INVENTION: AND FERREDOXIN-SOY IN
/ TITLE OF INVENTION: STREPTOMYCES
/ NUMBER OF SEQUENCES: 11
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: E. I. DU PONT DE NEMOURS
/ ADDRESSEE: AND COMPANY
/ STREET: 1007 MARKET STREET
/ CITY: WILMINGTON
/ STATE: DELAWARE
/ COUNTRY: USA
/ ZIP: 19898
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.50 inch,

```


CITY: Cleveland
STATE: Ohio
COUNTRY: USA
ZIP: 44114-2688
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/096,982
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Golrick, Mary E.
REGISTRATION NUMBER: 34829
REFERENCE/DOCKET NUMBER: 22727/0013
TELECOMMUNICATION INFORMATION:
TELEPHONE: 216-622-8458
TELEFAX: 216-241-0816
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2081 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 227..1649
US-09-096-982-7

Query Match: 9.2%; Score 46.2; DB 2; Length 2081;
Best Local Similarity 47.7%; Pred. No. 0.15;
Matches 135; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 108 CAACCGGCTGCTCCGAGGCTCCGCGCCCTTCCCTGGCGCGCGCGCGCGCGCG 167
DB 742 CGACGGGACGGATCGCATCGCGGCGCATCGCGGCGCTGCTCAGCGAATCGCGACTC 901
QY 168 GCGGCGCCCTCCGCGGAGTCCGCGCGCATAGCGGAACCTGACTATTTCTCTACGGGT 227
DB 802 CTCGACCGGTGCGGCGAACCGCGGAGCTGTCGCGGCTTCGCGTACACTTCGCGCT 861
QY 228 GTTGTAGGCTTCCGCGCTGCTGCTTGGTGGCTTCACCTGCGCGCTCTCTTCTGTTG 287
DB 862 GTTGTGATCTCGAACTGCTCGCGGTGCGCGGTACCGGATCGCGCAATGCGCGGAGGC 921
QY 288 GCTCTGCGAGGCTTCTCCGCGCGCTCATGCGAGCGCAAGAGAGGCTCCGCGCGCGCG 347
DB 922 CGTCGGCGGTGCTCAAGGCACTCGCGCTCGCGCGCGCGCGAGAGCGCGCGGTGACGCG 981
QY 348 AGCACTGCTGCTCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 390
DB 982 GGACCTGCGCGGACGTGCGCGGACACGTGCGCGCGCTGGAGAGC 1024

RESULT 8

US-09-096-982-7
Sequence 7, Application US/0969827
Patent No. 5976830
GENERAL INFORMATION:
APPLICANT: Strohl, William R.
APPLICANT: Dickens, Michael L.
APPLICANT: Desanti, Charles L.
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: CALFEE, HALTER & GRISWOLD
STREET: 800 Superior Avenue, Suite 1400
CITY: Cleveland
STATE: Ohio
COUNTRY: USA

ZIP: 44114-2688
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/653,650A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Golrick, Mary E.
REGISTRATION NUMBER: 34829
REFERENCE/DOCKET NUMBER: 22727/0013
TELECOMMUNICATION INFORMATION:
TELEPHONE: 216-622-8458
TELEFAX: 216-241-0816
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2081 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 227..1649
US-09-653-650A-7

Query Match: 9.2%; Score 46.2; DB 2; Length 2081;
Best Local Similarity 47.7%; Pred. No. 0.15;
Matches 135; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 108 CAACCGGCTGCTCCGAGGCTCCGCGCCCTTCCCTGGCGCGCGCGCGCGCGCG 167
DB 742 CGACGGGACGGATCGCATCGCGGCGCATCGCGGCGCTGCTCAGCGAATCGCGACTC 801
QY 168 GCGGCGCCCTCCGCGGAGTCCGCGCGCATAGCGGAACCTGACTATTTCTCTACGGGT 227
DB 802 CTCGACCGGTGCGGCGAACCGCGGAGCTGTCGCGGCTTCGCGTACACTTCGCGCT 861
QY 228 GTTGTAGGCTTCCGCGCTGCTGCTTGGTGGCTTCACCTGCGCGCTCTCTTCTGTTG 287
DB 862 GTTGTGATCTCGAACTGCTCGCGGTGCGCGGTACCGGATCGCGCAATGCGCGGAGGC 921
QY 288 GCTCTGCGAGGCTTCTCCGCGCGCTCATGCGAGCGCAAGAGAGGCTCCGCGCGCGCG 347
DB 922 CGTCGGCGGTGCTCAAGGCACTCGCGCTCGCGCGCGCGAGAGCGCGCGGTGACGCG 981
QY 348 AGCACTGCTGCTCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 390
DB 982 GGACCTGCGCGGACGTGCGCGGACACGTGCGCGCGCTGGAGAGC 1024

RESULT 9

US-09-096-982-6
Sequence 6, Application US/09096982
Patent No. 5962293
GENERAL INFORMATION:
APPLICANT: Strohl, William R.
APPLICANT: Dickens, Michael L.
APPLICANT: Desanti, Charles L.
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: CALFEE, HALTER & GRISWOLD
STREET: 800 Superior Avenue, Suite 1400
CITY: Cleveland
STATE: Ohio
COUNTRY: USA
ZIP: 44114-2688
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/096,982
FILING DATE:

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Goirick, Mary E.
REGISTRATION NUMBER: 34829
REFERENCE/DOCKET NUMBER: 22727/00131
TELECOMMUNICATION INFORMATION:
TELEPHONE: 216-622-8458
TELEFAX: 216-241-0816

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3013 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid

US-09-096-982-6
Query Match 9.2%; Score 46.2; DB 2; Length 3013;
Best Local Similarity 47.7%; Pred. No. 0.16;
Matches 135; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 108 CAACCGCGTGCCTCCAGGCGCTCCGCGCCCTTACCTGGGCGCGCGCGCGCGCGCGG 167
DB 1674 CGAGCGGACGGATCGATCGCGCGCATCGCGCGCATCGCGCGCATCGCGCGCATC 1733
QY 168 GCGGCG 227
DB 1734 CTCCGACCGGTGCGGCGGAAACCGGCGGAGCTGATCGCGCGGTGCGGCGGTGCGG 1793
QY 228 GTTGTAGGCTTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 287
DB 1794 GTTGTATCTGGAAGTCTGCGGCGTGGCGGCGTGGCGGCGTGGCGGCGTGGCGG 1853
QY 288 GCTGTGCGAGCGCTTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 347
DB 1854 CGTGGCGTGTCAAGGCACTCGCGCGTGGCGGCGTGGCGGCGTGGCGGCGTGGCG 1913
QY 348 AGCACCTGCCTCG 390
DB 1914 GGACCTGCGCGGCGACGTGCGCGGACACGTGCGCGGCTGGAGAGC 1956

RESULT 10
US-08-653-650A-6
Sequence 6, Application US/08653650A
Patent No. 5976830
GENERAL INFORMATION:
APPLICANT: Strohl, William R.
APPLICANT: Dickens, Michael L.
APPLICANT: DeSanti, Charles L.
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: CALFEE, HALTER & GRISWOLD
STREET: 800 Superior Avenue, Suite 1400
CITY: Cleveland
STATE: Ohio
COUNTRY: USA
ZIP: 44114-2689
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/653,650A
FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Goirick, Mary E.
REGISTRATION NUMBER: 34829
REFERENCE/DOCKET NUMBER: 22727/00131
TELECOMMUNICATION INFORMATION:
TELEPHONE: 216-622-8458
TELEFAX: 216-241-0816
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3013 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid

US-08-653-650A-6
Query Match 9.2%; Score 46.2; DB 2; Length 3013;
Best Local Similarity 47.7%; Pred. No. 0.16;
Matches 135; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 108 CAACCGCGTGCCTCCAGGCGCTCCGCGCCCTTACCTGGGCGCGCGCGCGCGCGG 167
DB 1674 CGAGCGGACGGATCGATCGCGCGCATCGCGCGCATCGCGCGCATCGCGCGCATC 1733
QY 168 GCGGCG 227
DB 1734 CTCCGACCGGTGCGGCGGAAACCGGCGGAGCTGATCGCGCGGTGCGGCGGTGCGG 1793
QY 228 GTTGTAGGCTTGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 287
DB 1794 GTTGTATCTGGAAGTCTGCGGCGTGGCGGCGTGGCGGCGTGGCGGCGTGGCGG 1853
QY 288 GCTGTGCGAGCGCTTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 347
DB 1854 CGTGGCGTGTCAAGGCACTCGCGCGTGGCGGCGTGGCGGCGTGGCGGCGTGGCG 1913
QY 348 AGCACCTGCCTCG 390
DB 1914 GGACCTGCGCGGCGACGTGCGCGGACACGTGCGCGGCTGGAGAGC 1956

RESULT 11
US-09-096-982-4
Sequence 4, Application US/09096982
Patent No. 5962291
GENERAL INFORMATION:
APPLICANT: Strohl, William R.
APPLICANT: Dickens, Michael L.
APPLICANT: DeSanti, Charles L.
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: CALFEE, HALTER & GRISWOLD
STREET: 800 Superior Avenue, Suite 1400
CITY: Cleveland
STATE: Ohio
COUNTRY: USA
ZIP: 44114-2688
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/096,982
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Goirick, Mary E.
REGISTRATION NUMBER: 34829
REFERENCE/DOCKET NUMBER: 22727/00131
TELECOMMUNICATION INFORMATION:

TELEPHONE: 216-622-8458
TELEFAX: 216-241-0816
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3196 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 1498..2764
FEATURE:
NAME/KEY: CDS
LOCATION: 1498..2764
US-09-096-982-4

Query Match 9.2%; Score 46.2; DB 2; Length 3196;
Best Local Similarity 47.7%; Pred. No. 3.16;
Matches 135; Conservative 0; Mismatches 148; Indels 0; Gaps 0;
QY 138 CAACCCGGTGCCTCCAGGCTCCGCCCCCTTGGCCCTGGCCCCGGCCCTTCCGGCCGGG 167
DB 1857 CGAGCGGACGGATCGCATCGCCGCCATCGCCGACCGGCTGTACCGAAGTCCGCCACTC 1916
QY 168 GCGGCCCCCTCCGCCCCGAGTCGCGGCATAGCGGACCTGTACTATTTTCCTACCCGGCT 227
DB 1917 CTCGACCGGTCCGCGCAACCGCGGACGTGCGGGGCTTCCGTACCACTTCCCGCT 1976
QY 228 GTTTGTAGGCTTCGGCTGCTGGGTTTGGTGGCTTCCAGTGGGGCTCTCTTGTGTG 287
DB 1977 GTTGTGATCTCGAAGTCTTGGCTGGCTGGCTGCGGTACCGATCGGCAATGCGCGGAGGC 2036
QY 288 GCTGTGCCAGCGCTTCTCCCGGCGCCCTCATGCGAGCCAGAGAGGCTCCGGGCGCGCGC 347
DB 2037 CGTCGGCGTGTTCAGGCACTCGGCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGAC 2096
QY 348 AGCACTGCTCGGCTCGGC 390
DB 2097 GGACCTGCGCGGACGTGCGCGGACAGTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGC 2139

RESULT 12
US-08-653-650A-4
Sequence 4, Application US/08653650A
Patent No. 5976930
GENERAL INFORMATION:
APPLICANT: Strohl, William R.
APPLICANT: Dickens, Michael L.
APPLICANT: Desanti, Charles L.
TITLE OF INVENTION: METHOD OF PRODUCING DOXORUBICIN
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: CALFEE, HALTER & GRISWOLD
STREET: 800 Superior Avenue, Suite 1400
City: Cleveland
State: Ohio
Country: USA
ZIP: 44114-2688
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/653.650A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Goldrick, Mary E.
REGISTRATION NUMBER: 34829
REFERENCE/DOCKET NUMBER: 22727/00131
TELECOMMUNICATION INFORMATION:

TELEPHONE: 216-622-8458
TELEFAX: 216-241-0816
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3196 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 1498..2764
FEATURE:
NAME/KEY: CDS
LOCATION: 1498..2764
US-08-653-650A-4

Query Match 9.2%; Score 46.2; DB 2; Length 3196;
Best Local Similarity 47.7%; Pred. No. 3.16;
Matches 135; Conservative 0; Mismatches 148; Indels 0; Gaps 0;
QY 108 CAACCCGGTGCCTCCAGGCTCCGCCCCCTTGGCCCTGGCCCCGGCCCTTCCGGCCGGG 167
DB 1857 CGAGCGGACGGATCGCATCGCCGCCATCGCCGACCGGCTGTACCGAAGTCCGCCACTC 1916
QY 168 GCGGCCCCCTCCGCCCCGAGTCGCGGCATAGCGGACCTGTACTATTTTCCTACCCGGCT 227
DB 1917 CTCGACCGGTCCGCGCAACCGCGGACGTGCGGGGCTTCCGTACCACTTCCCGCT 1976
QY 228 GTTTGTAGGCTTCGGCTGCTGGGTTTGGTGGCTTCCAGTGGGGCTCTCTTGTGTG 287
DB 1977 GTTGTGATCTCGAAGTCTTGGCTGGCTGGCTGCGGTACCGATCGGCAATGCGCGGAGGC 2036
QY 288 GCTGTGCCAGCGCTTCTCCCGGCGCCCTCATGCGAGCCAGAGAGGCTCCGGGCGCGCGC 347
DB 2037 CGTCGGCGTGTTCAGGCACTCGGCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGAC 2096
QY 348 AGCACTGCTCGGCTCGGC 390
DB 2097 GGACCTGCGCGGACGTGCGCGGACAGTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGC 2139

RESULT 13
US-07-928-611-12
Sequence 12, Application US/07928611
Patent No. 5569601
GENERAL INFORMATION:
APPLICANT: Van Tol, Hubert H.M.
APPLICANT: Civelini, Olivier
TITLE OF INVENTION: A No. 5569601 Human Dopamine Receptor and Uses
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Allegretti & Witcoff, Ltd.
STREET: 10 South Wacker Drive, Suite 3000
City: Chicago
State: Illinois
Country: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/928.611
FILING DATE: 19920810
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5569601nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 90,1092-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000


```
Matches 181; Conservative 0; Mismatches 217; Indels 0; Gaps 0;
QY 2 GTTCCCGTCGGTCTCGGGAGCGGGTTATGGCGGGCGGCGAGTGAAGCTGTGATGA 61
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
59 GTGCTCGGCTCAGCGGGGCTAGAAAGGCGCGGGACGGGCTGGAGTGGAGGGGGG 127
QY 62 ATTCTCCGGTGGACGAGGGAAGAAAGCTCCGGGAGCGCCAGACCAACCGGTGCTC 121
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
129 GAAGGGCGGACAGGGCGGGGCGGACGCTCTCTCGGCGAGGCTCAGCGGCGGGCC 187
QY 122 CCAGGCTCCGCGCCCTTTCGCTGGCCCGCCGCTCCCGGCGCGGCGCGCCCTCGGC 181
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
188 TCAGTCCGCGTCCGCGCTCGGCGCGCGCGCGCTAGCATGACGAGCTGTGTGGCG 247
QY 182 CCGAGTCCGCGCATAGCGGACCTGTACTATTTCCTACCGCGTGTGTGTAGGCTCG 241
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
248 CGGCCCCCAGCGCTGGAGAGGAGAACGACGGCTACTTTCGGAAGTGGCGGCTCG 307
QY 242 CGCTGCTCGCTTGGTGGCTTCCACCTGGGGCTCCTCTCTGTGTGTGTGTGTGT 301
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
308 GCGGGCGGGGCGGACCTGGGGCTCCAGACCTCTTTCGCGCGGCGCTCAGAGGAC 367
QY 302 TCTCCCGGCGCTCATGGCAGCAAGAGAGCTCCGGGCGCGCGGCGCGCTGCTG 361
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
368 GTGGAGCGCGAGTCCGGGTCCACTAGCGCGGAGGCGGAGGCGGCGACTGGCGCG 427
QY 362 CCTCGGCGCGCGCGCGCTGCGCGCGCGCGCGCGAGCGAGCGAG 399
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
428 GCTGGCGGAGCGAGGCTGGCGGAGGAGTGGGGGCG 465
```

RESULT 2

```
US-10-027-632-31508
; Sequence 31508, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; POLYMORPHISMS IN THE HUMAN GENOME
```

```
; FILE REFERENCE: 10827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/213,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
```

SEQ ID NO 31508

```
; LENGTH: 956
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(956)
; OTHER INFORMATION: n = A,T,C or G
```

US-10-027-632-31508

```
Query Match 9.6%; Score 48.2; DB 13; Length 956;
Best Local Similarity 46.2%; Pred. No. 0.0033;
Matches 146; Conservative 0; Mismatches 169; Indels 0; Gaps 0.
QY 9: GGTCGGCGGCGGCGGAGCAACCGGTGCTCCGAGCGGCTCGGCGCGGCGGCGGCG 150
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
DB 222 GGGCCCTGCCCCCTGCCCCCTGCCCCGCCCCGCCCCCCCCCCCCCCCCCCCCCCC 281
QY 151 GCGCCCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 210
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
282 CCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 341
QY 211 TATTCTCTACCGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 270
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
342 CCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 401
QY 271 GGGCTCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 320
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
402 CCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 461
QY 331 AGCTCCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 390
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
462 CCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 521
QY 391 GAGCGCGAGCGCGTCC 426
DB ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
522 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 581
```

RESULT 3

```
US-10-017-161-1403/c
; Sequence 1403, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
```

```
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/C152
; CURRENT APPLICATION NUMBER: US/10/017,161
; CURRENT FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
```

SEQ ID NO 1403

```
; LENGTH: 1117
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(1117)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(917)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (170)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (173)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (207)..(306)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (343)..(344)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (351)..(356)
; OTHER INFORMATION: a, t, c, g, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (360)..(361)
```

```
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (552)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (557)..(561)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (565)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (573)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (577)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (588)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (605)..(610)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (613)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (617)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (622)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (629)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (646)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (648)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (650)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (652)..(653)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (669)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (673)..(675)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (688)..(689)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:

OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (378)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (380)..(383)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (394)..(399)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (401)..(402)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (405)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (431)..(432)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (434)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (437)..(438)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (440)..(444)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (464)..(465)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (468)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (484)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (487)..(497)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (509)..(510)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (525)..(527)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (532)..(533)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (539)..(542)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (548)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
```



```

QY 288 GCTGTGCCAGCGCTTCTCCGGCGGCTCATGCGAGGCGAGAGAGAGCTCGGCGGCGCGCG 347
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 390 SAYSTSSSSSSSSSYVTSTNYC.T.CC...T.MCAABCSTTTTTTTTTT.HSCC.S 331

QY 348 AGCAGCTGCGCTCGGCTCGGCGCGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCG 407
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 330 A..A..M..YC.A.SYSYSYS.SSS.S.SYMR.HRA.SHYYTRS..S.MYCY.YM.Y..YY 271

QY 408 AGTCTTCCA 416
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 270 .YYSYVCSR 262

```

RESULT 6

```

US-10-142-885-10/c
; Sequence 10, Application US/10142885
; Publication No. US20030157604A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin

```

```

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P330R1C248

```

```

; CURRENT APPLICATION NUMBER: US/10/142,885

```

```

; CURRENT FILING DATE: 2002-05-10

```

```

; Prior Application removed - See File Wrapper or Palm

```

```

; NUMBER OF SEQ ID NOS: 560

```

```

; SEQ ID NO 10

```

```

; LENGTH: 594

```

```

; TYPE: PRT

```

```

; ORGANISM: Homo Sapien

```

```

US-10-142-885-10

```

```

Query Match          9.4%; Score 47; DB 12; Length 594;
Best Local Similarity 6.8%; Pred. No. 0.0067;
Matches 21; Conservative 152; Mismatches 132; Indels 4; Gaps 1;

```

```

QY 108 CAACCGGTCCTCCGCGAGCTCGCGGCTCGGCGGCTCGGCGGCGGCGGCGGCGGCGG 167
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 566 CA.MCCT..AMM.M.CBT.STT.VA.M.YT.S.S.S.SYSYSYS.S.S.SSYSYA.S 507

QY 168 GCCGGCCCTCCGCGCGAGTCGCGCGCATAGCGGAGCCTGTACTATTCTCCTACCGGCT 227
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 506 YSYS.S.S.SSYSYSSSDDY.CYCCYRYHCHSDSYSYSY.C---RCYVT.SYSRYD 451

QY 228 GTTGTAGGCTTCGCTGCGTTCGCTGGCTTCACCTCCAGCTGTACTATTCTCCTACCGGCT 287
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 450 CHYSCCCSDYYCYYSYRYYSYSYSWSYSYSYTDYCSYRCCCYYSYSSSYSS 391

QY 288 GCTCTGCCAGCGCTTCCTCCCGGCGGCTCATGGCGAGCGGCGGCGGCGGCGGCGGCGG 347
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 390 SAYSTSSSSSSSSSYVTSTNYC.T.CC...T.MCAABCSTTTTTTTTTT.HSCC.S 331

QY 348 AGCACCTGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 407
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 330 A..A..M..YC.A.SYSYSYS.SSS.S.SYMR.HRA.SHYYTRS..S.MYCY.YM.Y..YY 271

```

```

QY 408 AGTCTTCCA 416
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 270 .YYSYVCSR 262

```

RESULT 7

```

US-10-158-790-10/c
; Sequence 10, Application US/10158790
; Publication No. US20030180879A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin

```

```

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P330R1C448

```

```

; CURRENT APPLICATION NUMBER: US/10/158,790

```

```

; CURRENT FILING DATE: 2002-05-30

```

```

; Prior Application removed - See File Wrapper or Palm

```

```

; NUMBER OF SEQ ID NOS: 550

```

```

; SEQ ID NO 10

```

```

; LENGTH: 594

```

```

; TYPE: PRT

```

```

; ORGANISM: Homo Sapien

```

```

US-10-158-790-10

```

```

Query Match          9.4%; Score 47; DB 12; Length 594;
Best Local Similarity 6.8%; Pred. No. 0.0067;
Matches 21; Conservative 152; Mismatches 132; Indels 4; Gaps 1;

```

```

QY 108 CAACCGGTCCTCCGCGAGCTCGCGGCTCGGCGGCTCGGCGGCGGCGGCGGCGGCGG 167
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 566 CA.MCCT..AMM.M.CBT.STT.VA.M.YT.S.S.S.SYSYSYS.S.S.SSYSYA.S 507

QY 168 GCCGGCCCTCCGCGCGAGTCGCGCGCATAGCGGAGCCTGTACTATTCTCCTACCGGCT 227
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 506 YSYS.S.S.SSYSYSSSDDY.CYCCYRYHCHSDSYSYSY.C---RCYVT.SYSRYD 451

QY 228 GTTGTAGGCTTCGCTGCGTTCGCTGGCTTCACCTCCAGCTGTACTATTCTCCTACCGGCT 287
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 450 CHYSCCCSDYYCYYSYRYYSYSYSWSYSYSYTDYCSYRCCCYYSYSSSYSS 391

QY 288 GCTCTGCCAGCGCTTCCTCCCGGCGGCTCATGGCGAGCGGCGGCGGCGGCGGCGGCGG 347
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 390 SAYSTSSSSSSSSSYVTSTNYC.T.CC...T.MCAABCSTTTTTTTTTT.HSCC.S 331

QY 348 AGCACCTGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 407
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 330 A..A..M..YC.A.SYSYSYS.SSS.S.SYMR.HRA.SHYYTRS..S.MYCY.YM.Y..YY 271

QY 408 AGTCTTCCA 416
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 270 .YYSYVCSR 262

```

RESULT 8

```

US-10-123-155-10/c

```

```

; Sequence 10, Application US/10123155

```

```

; Publication No. US20030068794A1

```



```
LOCATION: source
FEATURE:
LOCATION: (1)...(3133)
FEATURE:
NAME/KEY: modified_base
LOCATION: (229)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (231)...(2933)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1)...(1115)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (124)...(129)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (139)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (146)...(148)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (151)...(152)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (155)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (172)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (179)...(180)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (182)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (185)...(188)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (192)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (194)...(195)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (210)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (212)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (214)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (219)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (222)...(225)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (229)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (231)...(232)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (238)...(239)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (242)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (245)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (248)...(260)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (265)...(270)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (272)...(276)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (278)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (282)...(283)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (287)...(290)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (302)...(306)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (311)...(313)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (323)...(334)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (336)...(341)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (345)...(363)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (365)...(368)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (370)
OTHER INFORMATION: a, t, c, g, unknown or other
```



```
RESULT 13
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US2003019018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, YASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 9.31; Score 46.4; DB 14; Length 9025608;
Best Local Similarity 52.0%; Pred. No. 0.01;
Matches 104; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 257 TCGCCCTCCACGGGGGGGCTCTTCCTGCTGGCTGTGCGAGCGCTTCTGCGCGCGCGCTCA 316
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 6193898 TCGCGCGCGCATCGCGGTCTCGCGCGGGGGCTCGCGGTGCGACGCGAGTGGCGCGGAGA 6194057
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 317 TGGCAGCCAGAGGAGGTCCGGGGCGGGCGCGGCGGCGGCTGCTGCGCTGGCGCGCGCGG 376
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 6193958 TCGCGCGCGCGGAGGGCGGTGGGGCGGCTGCTGCGCGCGCGGATCGGACAGCGCGCGG 6194017
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 377 CGGTGCGCGGGGGGAGGCGCGGCGGCGGCTGCTGCGAGCTTCTGCGACAGAGCGGCTTGGAGTACA 436
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 6194018 CCGCGGTGCGCGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 6194077
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 437 TCTCATTTGCGCTGGCATC 456
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 6194078 ACCGGAACGGCGGACGAGG 6194097
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
US-10-224-260-12
; Sequence 12, Application US/10224260
; Publication No. US20030059845A1
; GENERAL INFORMATION:
; APPLICANT: Van Tol, Hubert H.M.
; APPLICANT: Civelini, Olivier
; TITLE OF INVENTION: A No. US20030059845A1: Human Dopamine Receptor and Uses
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Allegretti & Witcoff, Ltd.
; STREET: 10 South Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/224,260
FILING DATE: 20-Aug-2002
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/928,611
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: No US20030059845A1:ran, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 95,1092-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
TELEX: 810-221-8317
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 803 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: exon
LOCATION: 1..803
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /evidence= EXPERIMENTAL
/standard name= "Alternate Exon 3: D4.7"
/note= "This sequence represents the third exon of
allele D4.7 of the human D4 dopamine receptor
Gene"
FEATURE:
NAME/KEY: misc feature
LOCATION: 257..262
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /function= "PstI site"
/evidence= EXPERIMENTAL
/standard name= "PstI site"
/label= PstI
/note= "This sequence is a PstI site whereby
digestion of human genomic DNA produces a RFLP"
FEATURE:
NAME/KEY: repeat region
LOCATION: 346..352
IDENTIFICATION METHOD: experimental
OTHER INFORMATION: /rpt_type= "tandem"
/evidence= EXPERIMENTAL
/rpt_unit= 346..354
/note= "This sequence is a repeat found in 7 known
alleles of the human D4 dopamine receptor gene
encoding a 16 amino acid sequence repeated 7 times"
FEATURE:
NAME/KEY: CDS
LOCATION: 2..803
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-10-224-260-12

Query Match 9.23; Score 46; DB 14; Length 803;
Best Local Similarity 49.7%; Pred. No. 0.012;
Matches 145; Conservative 0; Mismatches 145; Indels 2; Gaps 1;

QY 117 GCGTCCGAGCGCTCGCGCGCGCTTGGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 176
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 297 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 356
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 177 TCGCGCGCGAGTGGCGCGCATAGCGGAGCTGTACTATTCTCTACCGCGCTGTTGTAGG 236
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 357 GCGTCCCGCAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 416
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 237 CTTCGCGGTGCTGCGGTTTGGTCGCGCTTCCACCTGGGCGCTCTCTTCTGCTGTGGCTCTGCCA 296
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 417 GTCCCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 476
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```


FEATURES	Location/Qualifiers	
	source	
Tissue Procurement: Dr. David Rowe		
cDNA Library Preparation: Invitrogen Corp		
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)		
DNA Sequencing by: Agencourt Bioscience Corporation		
Clone distribution: MGC clone distribution information can be		
found through the I.M.A.G.E. Consortium/LLNL at:		
http://image.llnl.gov		
Plate: LLAX-4094 row: 3 column: 23		
High quality sequence start: 17		
High quality sequence stop: 520		
FEATURES		
Location/Qualifiers		
source		
1. 1106		
/organism="Mus musculus"		
/mol_type="mRNA"		
/db_xref="taxon:10090"		
/clone="IMAGE:6411742"		
/tissue_type="tumor, biopsy sample"		
/dev_stage="5 months"		
/lab_host="DH10B"		
/clone_lib="NCI CGAP Ma2"		
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;		
Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.		
Library constructed by Life Technologies. Investigator		
providing samples: Gilbert Smith, NIH"		
BASE COUNT	161 a 347 c 272 g 136 t	
ORIGIN		
Query Match 68.7%; Score 344; DB 13; Length 916;		
Best Local Similarity 86.5%; Pred. No. 4.3e-59;		
Matches 406; Conservative 0; Mismatches 60; Indels 6; Gaps 2;		
Qy	10 CGGTCTGCGGAGCGCGTTATGGCGGCGCGGCGCATGAGAGTGTGAAATGATTCGCG 63	
Db	1 CGGTCTGCGGAGCGCGTTATGGCGGCGCGGCGCATGAGAGTGTGAAATGATTCGCG 60	
Qy	70 GGTGGACGAGGAGAGAGAGAGCTCCGCGGCGCGGCGAGCAACCGGTCCCTCCAGGCGT 129	
Db	61 GCCGACGACGAGAGAGAGAGAGCTCCGCGGCGCGGCGAG---CCGCGCGCGCGAGGCGT 117	
Qy	130 CCGCGCGCGTTCCTGGCG 189	
Db	118 CCGCGCGCGCGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 174	
Qy	130 CCGCATAGCGGAGACCTGTACTATTCTCTAGCGCGCGCGCGCGCGCGCGCGCGCGCG 249	
Db	175 CCGCCTAGCGGAGACCGCT 214	
Qy	250 GGTTCGTGGCTTCACCTGGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 309	
Db	235 CGCCTGTGGCTTCACCTGGCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 294	
Qy	310 GCCCTCATGCGCAGCGAGAGCTCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 369	
Db	295 GCCCTCATGCGCAGCGAGAGCTCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 354	
Qy	370 CCGCGCGCGTTCGCGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 429	
Db	355 CCAGCGCGCGGACCGGTGGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 414	
Qy	430 GAGTACATTCGATTCG 491	
Db	415 GAGTACATTCGATTCG 466	
RESULT 2		
BUS18568	1106 bp mRNA linear EST 12-SEP-2002	
LOCUS	AGENCOURT 10164034 NIH MGC_134 Mus musculus cDNA clone	
DEFINITION	IMAGE:65163:0 5', mRNA sequence.	
BUS18568		
ACCESSION	BUS18568	
VERSION	BUS18568.1 GI:22826094	
KEYWORDS	EST.	
SOURCE	Mus musculus (house mouse)	
ORGANISM:	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Musinae; Mus.	
TITLE	1 (bases 1 to 1106)	
JOURNAL	NIH-MGC http://mgi.nci.nih.gov/	
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)	
	Unpublished	
	Contact: Robert Strausberg, Ph.D.	
	Email: cgapbs-r@mail.nih.gov	
BASE COUNT 182 a 379 c 325 g 220 t		
ORIGIN		
Query Match 56.1%; Score 281.2; DB 13; Length 1106;		
Best Local Similarity 84.7%; Pred. No. 1.6e-46;		
Matches 364; Conservative 0; Mismatches 58; Indels 8; Gaps 4;		
Qy	1 GGTTCGCGTGGTCTGGCGGAGCGGTTATGGCGGCGCGGCGAGTGAAGTGTGAATG 60	
Db	158 GGTTCGCGTGGTCTGGCGGAGCGGTTATGGCGGCGCGGCGAGTGAAGTGTGAATG 217	
Qy	61 AATTCTCGGGTGGAGGAGGAGAAAGAGTCTCGGCGCGCGCGCGCGCGCGCGCGCGCG 120	
Db	218 AGTTCTCGGGCGGACGACGAGGAGAAAGAGTCTCGGCGCGCGCGCGCGCGCGCGCG 274	
Qy	121 CCGAGGCGCTCGCGCGCGCGTGGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180	
Db	275 GCGAGGCGCTCGCGCGCGCGTGGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 331	
Qy	181 CCGAGTCTCGCGCGCGATAAGCGGAACTGTACTATTCTCTACCGCGCTGTGTAGGCTTC 240	
Db	332 GCGGCTCGCGCGCGTAAAGCGGAACTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 391	
Qy	241 GCGTCTCGTGGTGGTCTGCTGCTTCCACCTGCGGCGCTCTCTCTCTCTCTCTCTCT 300	
Db	392 GCGTCTCGTGGTGGTCTGCTGCTTCCACCTGCGGCGCTCTCTCTCTCTCTCTCTCT 451	
Qy	301 TTTTCGCGCGCGCTCATGGCAGCGGAGAGTCTCGGCGCGCGCGCGCGCGCGCGCGCG 360	
Db	452 TTTTCGCGCGCGCTCATGGCAGCGGAGAGTCTCGGCGCGCGCGCGCGCGCGCGCGCG 511	
Qy	361 GCGTCG 418	
Db	512 CCGTCG 571	
Qy	419 AACAGGCGCTT 428	
Db	572 ATCAGGCGCTT 581	
RESULT 3		
AQ937943	517 bp DNA linear GSS 23-AUG-2000	
LOCUS	AK937943	
DEFINITION	KB6-5799 Human NctI clones Homo sapiens genomic, genomic survey	
ACCESSION	AQ937943	
VERSION	AQ937943.1 GI:17214321	
KEYWORDS	GSS.	
SOURCE	Homo sapiens (human)	
ORGANISM	Homo sapiens	


```
OY 171 GGGCCCTCGGCGGAGTGGCGGATNAGCGGAACCTGTACTATTCTCTACAGCGCTGTT 230
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||
DB 1 GGGCCCTCGGCGGAGTGGCGGAGTAAAGGGAACCGCTCTCTCTACAGCGCTGTT 60
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||

OY 231 TGTAGGCTTCGGCTGCTGGCTTGGCTGGCTTCACCTCGGCGCTCTCTTTGTGTGGT 290
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||
DB 61 CGTCGGCTTCGGCTGCTGGCTGCTGGCTGGCTTCACCTCGGCGCTCTCTTTGTGTGGT 120
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||

OY 291 CTGGCAGGCTTCGGCTGCTGGCTGCTGGCTTCACCTCGGCGCTCTCTTTGTGTGGT 350
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||
DB 121 CTGGCAGGCTTCGGCTGCTGGCTGCTGGCTTCACCTCGGCGCTCTCTTTGTGTGGT 180
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||

OY 351 ACCTGGCTTCGGCTGCTGGCTGCTGGCTTCACCTCGGCGCTCTCTTTGTGTGGT 410
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||
DB 181 GCGCGCTTCGGCTGCTGGCTGCTGGCTTCACCTCGGCGCTCTCTTTGTGTGGT 240
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||

OY 411 CTTCACAAACAGGCTTCAGTACATCTCCATTGCTTGGCTTGGCTTGGCTTGGCTTGGT 470
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||
DB 241 CTTCACAAACAGGCTTCAGTACATCTCCATTGCTTGGCTTGGCTTGGCTTGGCTTGGT 300
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||

OY 471 AG 472
||
DB 301 AG 302

RESULT 6
BI987009
LOCUS
DEFINITION
3184-70 Mouse E14.5 retina lambda ZAP II library Mus musculus cDNA,
mRNA sequence.
ACCESSION
BI987009
VERSION
BI987009.1 GI:17957970
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 600)
Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,
White, R.A., Beremand, P.D., Thomas, T.J., Ganib, and Klein, W.H.
Gene expression in the developing mouse retina by EST sequencing
and microarray analysis
Nucleic Acids Res. 29 (24): 4963-4993 (2001)
MEDLINE
21671825
PubMed
11812828
Contact: Klein WH
Department of Biochemistry and Molecular Biology
University of Texas M.D. Anderson Cancer Center
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 792 8329.
FEATURES
source
location/Qualifiers
1..600
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/tissue_type="neural retina"
/dev_stage="embryonic day 14.5 post-fertilization"
/clone_lib="Mouse E14.5 retina lambda ZAP II library"
BASE COUNT
144 a 176 c 161 g 119 t
ORIGIN
Query Match 44.9%; Score 224.6; E12: Length 600;
Best Local Similarity 87.0%; Pred. No. 2.9e-35;
Matches 247; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

OY 198 GCGGAACCTGTACTATTCTCTACAGCGCTGTTGTAGGCTTCGGCTGCTGCTTGGT 257
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||
DB 1 GCGGAACCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 60
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||

OY 258 CGCCTTCACCTGGGCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 317
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||
DB 61 GGCCTGCCACCTGGGCTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 120
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||
```

```
OY 318 GGCAGGCAAGAGGAGTCTCGGGGCGCGCCAGCAGTCTGCTCGGCTCGGCGCGCGGCC 377
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||
DB 121 GCGCGCAAGAGGAGTCTCGGGGCGCGCCAGCAGTCTGCTCGGCTCGGCGCGCGGCC 180
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||

OY 375 GGTGCGGCGCGGAGGCGGAGCGGTCCGAGTCTCCACAAACAGGCTTCGAGTACAT 432
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||
DB 181 CGGACCGGCTGGGAGGCGGAGGCTCCGCGTCTCCACAGGAGCGCTTCGAGTACAT 240
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||

OY 438 CTCATTGCTGCGGATCGATGAGGATGAGAAAGTAACTAGS 481
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||
DB 241 CTCATTGCTGCGGATCGAGGAGGAGAAAGACAGAGAGS 284
||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 ||||| 171 |||||
```

```
RESULT 7
BX506397
LOCUS
DEFINITION
BX506397.1 GI:18670439
557 bp mRNA linear EST 12-MAR-2002
314305.Y1 Melton Normalized Mixed Mouse Pancreas cDNA clone IMAGE:5944376 5' similar to TR:09JPR9 09JPR9
X1AALC83 PROTEIN. 1; mRNA sequence.
```

```
ACCESSION
BX506397
VERSION
BX506397.1
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 557)
```

```
REFERENCE
1 (bases 1 to 557)
AUTHORS
Melton, D., Brown, J., Kenty, G., Permut, A., Lee, C., Kaestner, K.,
Lemishka, I., Scearce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
Hallier, L., Warra, M., Pape, D., Wylie, T., Martin, J., Blaisdell, A.,
Schmitt, A., Theising, B., Ritter, E., Ronko, J., Bennett, C., Cardenas
, V., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T.,
Jackson, Y. and Bowers, Y.
```

```
TITLE
Endocrine Pancreas Consortium
JOURNAL
Unpublished
COMMENT
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1912
Fax: 617-495-8557
Email: dmelton@biocp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Juliana Brown
brown@fas.harvard.edu
MGI:2007249 This sequence now available from the IMAGE consortium,
for clone orders contact: info@image.llnl.gov
Seq primer: -4CRP from Gibco
High quality sequence stop: 435.
location/Qualifiers
1..557
/organism="Mus musculus"
/mol_type="mRNA"
/strain="ICR"
/db_xref="taxon:10090"
/clone="IMAGE:5944376"
/sex="Both for embryonic & newborn, male for adult and
adult islet"
/dev_stage="Embryonic day 10.5, E12.5, E16.5, newborn,
adult, mixed"
/lab_host="CH10E"
/clone_lib="Melton Normalized Mixed Mouse Pancreas 1
NI-MMS1"
```

```
FEATURES
source
location/Qualifiers
1..557
/organism="Mus musculus"
/mol_type="mRNA"
/strain="ICR"
/db_xref="taxon:10090"
/clone="IMAGE:5944376"
/sex="Both for embryonic & newborn, male for adult and
adult islet"
/dev_stage="Embryonic day 10.5, E12.5, E16.5, newborn,
adult, mixed"
/lab_host="CH10E"
/clone_lib="Melton Normalized Mixed Mouse Pancreas 1
NI-MMS1"
Note="Vector: pSPORT1; Site 1: Not 1; Site 2: Sal I; Five
libraries representing E10.5/12.5 pancreatic bud, E16.5
pancreas, newborn pancreas, adult pancreas, and adult
islets of Langerhans were separately constructed using
SuperScript Plasmid Library kit (Life Technologies). cDNA
was made by oligo-dT priming and size-selected by column
fractionation. Libraries were amplified once on solid
```



```
Db 182 GACCGGTGGCGGAGCGGAGAGCGTCCGNGTCTTCCACAGCAGGCGCTTCGAGTACAT 241
QY 442 CCATTGCGCTGCGCATCGATGAGGATGAGAGAGGTAAGTAGG 481
Db 242 CCATTGCGCTGCGCATCGAGGAGGAGAGAGAGGAGGAGG 283

RESULT 9
BG391372
LOCUS 50391372
DEFINITION 602417440P1 NIH_MGC_22 Homo sapiens cDNA clone IMAGE:4537184 5',
RNA sequence.
ACCESSION BG391372
VERSION BG391372.1 GI:13264520
KEYWORDS EST.
ORGANISM Homo sapiens (human);
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 976)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC);
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.V.A.G.E. Consortium (LINC)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.V.A.G.E. Consortium/LINC at:
http://image.lnl.gov
Plate: LLAM10461 row: 0 column: 09
High quality sequence stop: 725.
Location/Qualifiers
1..976
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4537184"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_92"
/Note="Organ: testis; Vector: pCMV-Sport6; Site 1: Not;
Site 2: Salt; Cloned unidirectionally; oligo-dT primed.
Average insert size 2.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC library."
BASE COUNT 287 a 237 c 253 g 199 t
ORIGIN

Query Match 43.1%; Score 216; DA 10; Length 976;
Best Local Similarity 99.2%; Pred. No. 1.7e-33;
Matches 238; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 233 TAGGCTTCGGCTGCTGCGCTTGGTCGGCTTCACCTGCGGCTTCGTCGCTCT 292
Db 1 TAGGCTTCGGCTGCTGCG-TTGGTCGGCTTCACCTGCGGCTTCGTCGCTCT 50
QY 293 GCCAGGCTTCCTCCCGCGGCTCATGGCAGCCCAAGAGAGTCCGGGCGCGCGGAGCAC 352
Db 60 GCCAGGCTTCCTCCCGCGGCTCATGGCAGCCCAAGAGAGTCCGGGCGCGCGGAGCAC 319
QY 353 CTGCTCTGGCTTCGGCCCGCGGCGCGGTCGGCGCGGCGGCGGCGGCGGCTCCAGTCT 412
Db 120 CTGCTCTGGCTTCGG-CCCGGCGCGGTCGGCGGCGGCGGCGGCGGCGGCTCCAGTCT 178
QY 413 TCCACAAACAGCGCTTCGAGTACATCTCCATTGCCCTGGCGCATCGATGAGGATGAGAAAG 472
Db 179 TCCACAAACAGCGCTTCGAGTACATCTCCATTGCCCTGGCGCATCGATGAGGATGAGAAAG 238
```

```
RESULT 10
BB618976
LOCUS 553 bp mRNA linear EST 26-OCT-2001
DEFINITION BB618976 RIKEN full-length enriched, 8 days embryo Mus musculus
cDNA clone 5730458M19 5', mRNA sequence.
ACCESSION BB618976
VERSION BB618976.1 GI:16458433
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 553)
AUTHORS Hiramoto, K., Carninci, P., Fukuda, S., Furuno, M., Hatanaka, T., Hata, A.,
M., Koya, S., Matsuyama, T., Miyazaki, A., Nomura, K., Ohno, M.,
Okazaki, Y., Okido, T., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki,
O., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H.,
Tagami, M., Tagawa, A., Takahashi, F., Takeda, Y., Tanaka, T., Toya, T.,
Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Arakawa, T., et al. 2001)
Unpublished
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh,
X., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura,
S., Kawai, S., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multichannel sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara,
Y. and Hayashizaki, Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamanaka, T., Aizawa,
K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and
Hayashizaki, Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.
Location/Qualifiers
1..553
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="5730458M19"
/sex="mixed"
/dev_stage="8 days embryo"
/lab_host="DH10B"
/clone_lib="RIKEN full-length enriched, 8 days embryo"
/Note="Site 1: Salt; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
```


Db 400 CGTCGGAGTC-CGGCATAGCGGAACTGTACTATTCTCTCTACCGGCTGTT--GTAGGC 456

QY 239 TCGCGCTGCTGGCTTTGGTGGCTTCAC 267
|||||
Db 457 TCTCGCTGCTGGCTTTGGTGGCTTCAC 485
|||||

RESULT 15
AQ683149 548 bp DNA linear GSS 28-JUN-1999
LOCUS HS_5378_32_c12_T7A_RP01-11 Human Male BAC Library Homo sapiens
DEFINITION genomic clone Plate=954 Col=24 Row=F, genomic survey sequence.
ACCESSION AQ683149
VERSION AQ683149.1 GI:5259055
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1. (bases 1 to 548)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
PUBMED 10449764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3897
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RP01-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
<http://www.htsc.washington.edu>
Plate: 954 row: F column: 24
Seq primer: T7
Class: BAC ends
High quality sequence stop: 548.
Location/Qualifiers
1. 548
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="Plate=954 Col=24 Row=F"
/sex="male"
/cclone_lib="RP01-11 Human Male BAC Library"
/note="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI MethyIase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"
BASE COUNT 128 a 159 c 142 g 115 t 4 others
ORIGIN

Query Match 37.2%; Score 186.6; DS 28; Length 548;
Best Local Similarity 66.4%; Pred. No. 1.2e-27;
Matches 267; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 100 GGCGCAGCAACCGGTCCTTCACAGGCTTCGCGCCCTTCCTGGCGCGGCGGCGGCTGCC 159
|||
Db 72 GGGATCATCGACCGCGGCACACATTTCTCTGCTTCATCTCTCTGAGCGGCGGCGGCTTA 131
|||

QY 160 GCGCGCGGCGGCGGCGGCTTCGCGCGGAGTCGCGCGCATAGCGGAACCTGACTATTCTCC 219
|||

Db 132 GATGTCGGCCAGTCAGTGTGNGCATATGGGCGCATAGCATACATAGGCTATTCTTAA 191
|||

QY 220 TACCGGCTGTTTGTAGGCTTGGCGCTGCTGGCTTTGGTCGCTTCCACCTGGGGCTCCTC 279
|||||
Db 192 CACCGGCTGCATGTAGCTTCGAGCTGATGGACCGAGTCCCTTCCAGCTGAGGACCTC 251
|||||

QY 280 TTGTGTGGCTCTGCCAGCGCTTCTCCCGCGCGGCTCATGGGAGCCCAAGAGGAGCTCCGGG 339
|||||
Db 252 TTGAGAGGCTTTGACATCGCTTCTCCGAAACCTCATGGGAGGCAATAGGAGCTCTCTGC 311
|||||

QY 340 GCGGCGCCAGCACTTCCCTCGGCTTCGGCTTCGGCTTCGGCTTCGGCTTCGGCTTCGGCTTC 389
|||||
Db 312 GCGACACCATCATCTGGCTAAGCTAGGGCGGCGGCGGCGGCGGCGGCTGATGCGGAC 371
|||||

QY 400 GCGCTCGGAGTCTTCCACAAACAGGCGCTTCGAGTACATCTGCTTCCCTTCGCTTCGAT 459
|||||
Db 372 CACGTCCAGTCTTCCACAAAGACTAGCGTTACAGTTTCATGTACATTCTCTGCGGCTCCCA 431
|||||

QY 460 GAGGATGAGAAAGGTAACTAGGCGGCTGGGCGGAGGGGCGGCGGC 501
|||||
Db 432 GAGGATGAGAAACAGCAACTAACCGGCTGGGCGGAGGCGGCGGCGGC 473
|||||

Search completed: October 24, 2003, 22:51:04
Job time : 1049.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

CV nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:43:05 : Search time 1381.12 seconds
(without alignments)
14808.112 Million cell updates/sec

Title: US-09-830-902-1_COPY_1_500
Perfect score: 500
Sequence: 1 taactgacctgtgtgtgtt.....aacatggtaaacaccatgttcg 500

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 20454913386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:

- 1: gb_ba:
- 2: gb_htg:
- 3: gb_in:
- 4: gb_om:
- 5: gb_ov:
- 6: gb_pat:
- 7: gb_ph:
- 8: gb_pl:
- 9: gb_pri:
- 10: gb_ro:
- 11: gb_sts:
- 12: gb_sy:
- 13: gb_un:
- 14: gb_vit:
- 15: em_ba:
- 16: em_fun:
- 17: em_hum:
- 18: em_in:
- 19: em_mu:
- 20: em_om:
- 21: em_or:
- 22: em_ov:
- 23: em_pat:
- 24: em_ph:
- 25: em_pi:
- 26: em_ro:
- 27: em_sts:
- 28: em_un:
- 29: em_vit:
- 30: em_htg_hum:
- 31: em_htg_inv:
- 32: em_htg_other:
- 33: em_htg_mus:
- 34: em_htg_pln:
- 35: em_htg_rod:
- 36: em_htg_mam:
- 37: em_htg_vrt:
- 38: em_sy:
- 39: em_htgo_hum:
- 40: em_htgo_mus:
- 41: em_htgo_other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	500	100.0	101584	9	CNS01D55	AL121655 BAC seque
2	500	100.0	110000	6	AX093471	AX093471 Sequence
3	500	100.0	110000	9	HS0246003	AJ246003 Homo sapi
4	500	100.0	155943	9	AC012364	AC012364 Homo sapi
C	203.8	40.8	133568	9	HS16915	Z93015 Human DNA s
C	180.8	38.2	117751	9	AC020913	AC020913 Homo sapi
C	189.6	38.0	45195	9	AC090421	AC090421 Homo sapi
9	189.6	37.9	156589	9	HS537823	AL034405 Human DNA
10	187.6	37.5	172837	9	AL450304	AL450304 Human DNA
11	181.4	36.3	169334	2	AC135837	AC135837 Papio anu
12	181.4	36.3	184341	2	AC135835	AC135835 Papio anu
13	179.4	35.9	89319	2	AL929255	AL929255 Homo sapi
C	179.4	35.9	97621	9	AL607089	AL607089 Human DNA
C	179.4	35.9	121362	9	AL359203	AL359203 Human DNA
15	179.4	35.9	145736	9	AL451139	AL451139 Human DNA
C	179.4	35.9	152428	9	AC114490	AC114490 Homo sapi
17	179.2	35.8	124337	9	HSBA436C9	AL121825 Human DNA
C	179	35.8	159130	9	AC001504	AC001504 Pan trogl
19	177.4	35.5	188546	2	AC022396	AC022396 Homo sapi
20	177.2	35.4	171309	9	AC012146	AC012146 Homo sapi
21	176.8	35.4	90214	9	AL356488	AL356488 Human DNA
22	176.8	35.4	190649	9	CNS01DX1	AL139317 Human chr
C	176.6	35.3	186902	9	AC010863	AC010863 Homo sapi
C	176.4	35.3	90143	9	AL669876	AL669876 Human DNA
C	176.4	35.3	151980	9	AC084209	AC084209 Homo sapi
25	176.4	35.2	106034	9	AC134028	AC134028 Homo sapi
26	176	35.2	154788	2	AC031984	AC031984 Homo sapi
C	176	35.2	186252	2	AC144518	AC144518 Pan trogl
C	175.8	35.2	110000	9	AE014305_0	AE014305 Homo sapi
C	175.8	35.2	164061	9	AL445188	AL445188 Human DNA
31	175.8	35.2	204439	9	AP000807	AP000807 Homo sapi
32	175.6	35.1	39729	6	AX647093	AX647093 Sequence
33	175.6	35.1	64475	9	AF283320S2	AF283321 Homo sapi
34	175.6	35.1	66933	6	AX277532	AX277532 Sequence
35	175.6	35.1	66933	6	AX418096	AX418096 Sequence
36	175.6	35.1	137206	2	AC024123	AC024123 Homo sapi
37	175.6	35.1	198720	9	AP002366	AP002366 Homo sapi
C	175.4	35.1	78167	9	AC092028	AC092028 Homo sapi
39	175	35.0	105087	9	AC093884	AC093884 Homo sapi
40	175	35.0	128152	9	AL136313	AL136313 Human DNA
41	174.8	35.0	59257	9	AC106019	AC106019 Homo sapi
42	174.8	35.0	110144	9	AC08053	AC08053 Homo sapi
C	174.8	35.0	130855	9	AC004089	AC004089 Homo sapi
C	174.8	35.0	188340	9	AC135178	AC135178 Homo sapi
45	174.8	35.0	189623	2	AC018640	AC018640 Homo sapi

ALIGNMENTS

RESULT 1
CNS01D55
LOCUS
DEFINITION BAC sequence from the SP04 candidate region at 2p21-2p22 BAC 336P14 of library CITB_978_SKB from chromosome 2 of Homo sapiens (Human).
ACCESSION AL121655
VERSION AL121655.1 GI:6002386
KEYWORDS SP04 genomic DNA interval.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 101584)
Hazan, J., Fonknechten, N., Mavei, D., Paternotte, C., Sanson, D.,
Artiguenave, F., Davoine, C.S., Cruaud, C., Durr, A., Wincker, P.,

Brottier, P., Cattolacci, L., Barbe, V., Burgunder, J.M.,
 Prud'Homme, J.F., Brice, A., Fontaine, B., Heilig, R. and
 Weissenbach, J.
 Spastin, a novel AAA protein, is altered in the most frequent form
 of autosomal dominant spastic paraplegia
 Nat. Genet. (1999) In press
 2 (bases : to 101584)
 Genoscope.
 Direct Submission
 Submitted (18-APR-2002) Genoscope - Centre National de Séquençage :
 BP 131 91006 EVRY cedex - FRANCE (E-mail : secre@genoscope.cns.fr)
 - Web : www.genoscope.cns.fr

```

FEATURES
source
1..101584
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/clone="336P14"
/clone_lib="CITA_978_SK3"
29860 a 22:12 c 20047 g 27567 t
BASE COUNT
ORIGIN

```

29860 a 22:10 c 22:47 g 27:56 f

QY	301	TCCTCCTGCTCAATCTTCCAAAGTAGCTGGGACTACAGGATATGAGCTGCGTACCTGGGAC	360
DB	301		
QY	361	TGGTTTGTGTTTTTAAATTTTGGAGCCAGGTGCAGTGGGCGATATTTGTGATCCGAGC	420
DB	361		
QY	421	ACTTTGGGAGACCAAGCGAGCGCGATTACTTGGAGGTCAGTGGAGTTCAGGACGACCGAGGC	480
DB	421		
QY	481	AACATGGTAAACCATGTGCG	500
DB	481		
DB	481	AACATGGTAAACCATGTGCG	500
DB	481		
RESULT 3			
HSA246003			
LOCUS		110000 bp	DNA linear PFI 09-MAR-2000
DEFINITION			Homo sapiens Spast gene for spastin protein.
ACCESSION			AC246003
VERSION			AC246003.1 GI:6273492
KEYWORDS			Spast gene; spastin protein; SPG4-linked hereditary spastic paraplegia.
SOURCE			Homo sapiens (human)
ORGANISM			Homo sapiens
REFERENCE			1
AUTHORS			Hazan, J., Fonknechten, N., Mavel, D., Paternotte, C., Sanson, D., Arriguenave, F., Davoine, C.S., Cruaud, C., Durr, A., Winkler, P., Brotier, P., Cattolico, L., Barbe, V., Burgunder, J.M., Prud'homme, J.F., Brice, A., Fontaine, B., Prud'homme, J., Weissenbach, J., Durr, A. and Hazan, J.
TITLE			Spastin, a new AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia
JOURNAL			Nat. Genet. 23 (3), 296-303 (1999)
MEDLINE			20055425
PUBMED			10610178
REFERENCE			2
AUTHORS			Fonknechten, N., Mavel, D., Byrne, P., Davoine, C., Cruaud, C., Boetsch, D., Sanson, D., Coutinho, P., Hutchinson, M., McMenagie, P., Burgunder, J., Tartaglione, A., Heitzel, O., Paki, I., Deufel, T., Parfrey, N., Brice, A., Fontaine, B., Prud'homme, J., Weissenbach, J., Durr, A. and Hazan, J.
TITLE			Spectrum of SPG4 mutations in autosomal dominant spastic paraplegia
JOURNAL			Hum. Mol. Genet. 9 (4), 637-644 (2000)
MEDLINE			20164302
PUBMED			10695187
REFERENCE			3
AUTHORS			Genoscope.
TITLE			Direct Submission
JOURNAL			Submitted (17-JUN-1999): Genoscope, Genoscope - Centre National de Sequencage, BP 191, EVRY 91006, FRANCE
COMMENT			E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr. The sequence is the result of the assembly of 2 PAC clones: R-336p14 and 563N4, respectively from RPCI-II and CITA_978_SKB library.
FEATURES			Location/Qualifiers
source			1..110000
			/organism="Homo sapiens"
			/mol_type="genomic DNA"
			/db_xref="taxon:9606"
			/chromosome="2"
			/map="2p21"
			/clone="R-336p14"
			/clone_lib="RPCI-II"
			/note="BAC R-336p14 contains Spast gene: exon 1 to exon 3, BAC 563N4 contains Spast gene: exon 4 to exon 17"
gene			9932..102009
			/gene="Spast"

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RP11-11 human BAC library was made from the blood of one male donor, as described by Choegawa, K., Woon, P.Y., Zhao, S., Prengen, B., Tateno, M., Catanese, D.C. and de Jong, P.J. (1995). An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-431P19; the clone sequenced to the right is RP11-78E13. Actual start of this clone is at base position 1 of RP11-444D15; actual end is at base position 155943 of RP11-444D15.

Data from AC011232 and AC010981 was used to finish this clone, AC012364. The sequence fidelity between bases 112626 to 112643 can not be guaranteed due to an unresolved homopolymeric run. The sequence between 112648 to 112663 is single stranded.

FEATURES

source

```

1..155943
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-444D15"
/clone_lib="RP11-1"
310..617
/rpt_family="Alu"
593..617
/rpt_family="AT-rich"
728..748
/rpt_family="AT-rich"
732..1026
/rpt_family="Alu"
1169..1242
/rpt_family="TATATATGin"
1610..2503
/rpt_family="C2"
2504..2776
/rpt_family="Alu"
2777..3498
/rpt_family="L2"
3590..3742
/rpt_family="MIR"
4452..4765
/rpt_family="Alu"
4550..4975
/note="match to EST AA632695 (NID:G2556109) np83b32.s1"
4661..4692

```

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

/note="match to EST BG5711163 (NID:G13578829)"

4970..5239

/rpt_family="Alu"

5211..5238

/rpt_family="CAAAAAln"

5234..5328

/note="match to EST BG5711163 (NID:G13578829)"

5242..5328

/note="match to EST BG5757662 (NID:G14066315)"

5247..5328

/note="match to EST AU121379 (NID:G10936614)"

5247..5328

/note="match to EST BP205430 (NID:G11099016)"

5247..5328

/note="match to EST BG0333257 (NID:G12425431)"

5247..5328

/note="similar to Mus musculus EST AA619874 (NID:G2523750)"

5247..5328

/note="similar to Mus musculus EST AW120623 (NID:G6095956)"

5740..6346

/rpt_family="L2"

6347..6439

/rpt_family="ACHobc"

6453..6777

/rpt_family="L2"

6938..7406

/rpt_family="L1"

7415..8075

/rpt_family="ERV1"

7900..8421

/note="match to EST AW994939 (NID:G8255173)"

8387..8245

/rpt_family="L1"

8447..8524

/rpt_family="MER2_type"

8525..8799

/rpt_family="Alu"

8800..9067

/rpt_family="MER2_type"

9917..10069

misc_feature

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

13820..14405

/note="match to EST BG434159 (NID:G13340653)"

14067..14097

/rpt_family="AT-rich"

14793..15032

/note="similar to Homo sapiens EST BG110235 (NID:G12604341)"

13820..14405

/note="match to EST BG434159 (NID:G13340653)"

14067..14097

/rpt_family="AT-rich"

14793..15032

/note="similar to Homo sapiens EST BG110235 (NID:G12604341)"

13820..14405

/note="match to EST BG434159 (NID:G13340653)"

14067..14097

/rpt_family="AT-rich"

14793..15032

/note="similar to Homo sapiens EST BG110235 (NID:G12604341)"

13820..14405

/note="match to EST BG434159 (NID:G13340653)"

14067..14097

/rpt_family="AT-rich"

14793..15032

/note="similar to Homo sapiens EST BG110235 (NID:G12604341)"

13820..14405

/note="match to EST BG434159 (NID:G13340653)"

14067..14097

/rpt_family="AT-rich"

14793..15032

/note="similar to Homo sapiens EST BG110235 (NID:G12604341)"

13820..14405

/note="match to EST BG434159 (NID:G13340653)"

14067..14097

/rpt_family="AT-rich"

14793..15032

/note="similar to Homo sapiens EST BG110235 (NID:G12604341)"

13820..14405

/note="match to EST BG434159 (NID:G13340653)"

14067..14097

/rpt_family="AT-rich"

Job time : 1386.52 secs

FT	exon	357493..358333
FT		/*tag= f
FT	intron	/number= "3"
FT		358333..450222
FT		/*tag= g
FT	exon	/number= "3"
FT		450223..451118
FT		/*tag= h
FT	intron	/number= "4"
FT		451119..608663
FT		/*tag= i
FT	exon	/number= "4"
FT		608664..610511
FT		/*tag= j
FT	intron	/number= "5"
FT		610512..619277
FT		/*tag= k
FT	exon	/number= "5"
FT		61928..620663
FT		/*tag= l
FT	intron	/number= "6"
FT		620662..623344
FT		/*tag= m
FT	exon	/number= "6"
FT		623345..624338
FT		/*tag= n
FT	intron	/number= "7"
FT		624339..732173
FT		/*tag= o
FT	exon	/number= "7"
FT		73174..73248
FT		/*tag= p
FT	intron	/number= "8"
FT		73249..746333
FT		/*tag= q
FT	exon	/number= "8"
FT		746334..74765
FT		/*tag= r
FT	intron	/number= "9"
FT		74766..82788
FT		/*tag= s
FT	exon	/number= "9"
FT		82789..82964
FT		/*tag= t
FT	intron	/number= "10"
FT		82965..83102
FT		/*tag= u
FT	exon	/number= "10"
FT		83103..83194
FT		/*tag= v
FT	intron	/number= "11"
FT		83195..83334
FT		/*tag= w
FT	exon	/number= "11"
FT		83335..83414
FT		/*tag= x
FT	intron	/number= "12"
FT		83415..89129
FT		/*tag= y
FT	exon	/number= "12"
FT		89130..88272
FT		/*tag= z
FT	intron	/number= "13"
FT		88173..89561
FT		/*tag= aa
FT	exon	/number= "13"
FT		89562..89641
FT		/*tag= ab
FT	intron	/number= "14"
FT		89642..91162
FT		/*tag= ac
FT	exon	/number= "14"
FT		91163..91233

FI		/ *tag= ad	
FI		/number= "15"	
FI	intron	91234..93443	
FI		/ *tag= ac	
FI		/number= "15"	
FI	exon	93444..93484	
FI		/ *tag= af	
FI		/number= "16"	
FI	intron	93485..100599	
FI		/ *tag= ag	
FI		/number= "16"	
FI	exon	100600..102009	
FI		/ *tag= ah	
FI		/number= "17"	
XX			
PK	FR2798138-A1.		
XX			
PD	09-MAR-2001.		
XX			
PF	03-SEP-1999;	99FR-0011097.	
XX			
PR	03-SEP-1999;	99FR-0011097.	
XX			
PA	(CNRS) CNRS CENT NAT RECH SCI.		
XX			
PI	Weissenbach J., Hazan J.		
XX			
DR	WP1; 2001-283966/3C.		
XX			
PT	New human nucleic acid from the		
PT	autosomal dominant familial spa		
XX			
PS	Claim 2; Page 45-106; 145pp; Fr		
XX			
CC	The present sequence represents		
CC	a spastin polypeptide. Mutation		
CC	autosomal dominant familial spa		
CC	and their fragments, are used to		
CC	encode spastin (particularly se		
CC	mice); to identify SPG4 mutatio		
CC	particularly for diagnosis of a		
CC	paraplegia (PSF-AD); to identif		
CC	of the SPG4 gene; for detection		
CC	production of spastin; and for		
XX			
SC	Sequence 110000 BP; 30622 A; 21		
	Query Match	100.0%;	
	Best Local Similarity	100.0%; P	
	Matches	500; Conservative	0;
QY	1	TAACGACCTCGCTGCTGTTTCG	
ED	1	TAACGACCTCGCTGCTGTTTCG	
QY	61	CAGTTTACTTACACCTACTTA	
ED	61	CAGTTTACTTACACCTACTTA	
QY	121	AACAAATCCTCCATAAAGTAGA	
ED	121	AACAAATCCTCCATAAAGTAGA	
QY	181	TTTGTGTTGTTTATTTTATTTT	
ED	181	TTTGTGTTGTTTATTTTATTTT	
QY	241	GCTGGAAATGCAATGGCATGATC	
ED	241	GCTGGAAATGCAATGGCATGATC	
QY	301	TCTCTGCGCTCAATCTTCCAA	

RESULT 7
AAK83537/c
ID AAK83537 standard; DNA; 314 BP.
XX
AC AAK83537;
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38349.
XX
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
QS Homo sapiens.
XX
PN WC200157182-A2.
XX
XX 09-AUG-2001.
PD
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-CAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184654.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227509.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232197.
PR 14-SEP-2000; 2000US-0232198.
PR 14-SEP-2000; 2000US-0232199.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234937.
PR 25-SEP-2000; 2000US-0234938.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237017.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.

RESULT 9	
AAK70270/c	
ID AAK70270 standard; DNA; 51469 BP.	
XX AC AAK70270;	
XX AC	
XX DT	
XX XX	06-NOV-2000 (first entry)
DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:25052.	
XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;	
KW cytostatic; gene therapy; vaccine; metastasis; ds.	
XX OS Homo sapiens.	
OS XX	
PN WO200157182-A2.	
XX PD	
XX XX	09-AUG-2001.
XX PF	
XX XX	17-JAN-2001; 2001WO-US01354.
PR 31-JAN-2000; 2000US-0179065.	
PR 04-FEB-2000; 2000US-0180628.	
PR 24-FEB-2000; 2000US-0184664.	
PR 02-MAR-2000; 2000US-0186350.	
PR 16-MAR-2000; 2000US-0189874.	
PR 17-MAR-2000; 2000US-0190076.	
PR 18-APR-2000; 2000US-0198123.	
PR 19-MAY-2000; 2000US-0205515.	
PR 07-JUN-2000; 2000US-0209467.	
PR 28-JUN-2000; 2000US-0214886.	
PR 30-JUN-2000; 2000US-0215135.	
PR 07-JUL-2000; 2000US-0216647.	
PR 07-JUL-2000; 2000US-0216880.	
PR 11-JUL-2000; 2000US-0217487.	
PR 11-JUL-2000; 2000US-0217496.	
PR 14-JUL-2000; 2000US-0218290.	
PR 26-JUL-2000; 2000US-0220963.	
PR 26-JUL-2000; 2000US-0220964.	
PR 14-AUG-2000; 2000US-0224518.	
PR 14-AUG-2000; 2000US-0224519.	
PR 14-AUG-2000; 2000US-0225213.	
PR 14-AUG-2000; 2000US-0225214.	
PR 14-AUG-2000; 2000US-0225266.	
PR 14-AUG-2000; 2000US-0225267.	
PR 14-AUG-2000; 2000US-0225268.	
PR 14-AUG-2000; 2000US-0225270.	
PR 14-AUG-2000; 2000US-0225447.	
PR 14-AUG-2000; 2000US-0225757.	
PR 14-AUG-2000; 2000US-0225758.	
PR 14-AUG-2000; 2000US-0225759.	
PR 18-AUG-2000; 2000US-0226279.	
PR 22-AUG-2000; 2000US-0226681.	
PR 22-AUG-2000; 2000US-0226688.	
PR 22-AUG-2000; 2000US-0227182.	
PR 23-AUG-2000; 2000US-0227009.	
PR 30-AUG-2000; 2000US-0228924.	
PR 01-SEP-2000; 2000US-0229287.	
PR 01-SEP-2000; 2000US-0229343.	
PR 01-SEP-2000; 2000US-0229344.	
PR 01-SEP-2000; 2000US-0229345.	
PR 05-SEP-2000; 2000US-0229509.	
PR 05-SEP-2000; 2000US-0229513.	
PR 06-SEP-2000; 2000US-0230437.	
PR 06-SEP-2000; 2000US-0230438.	
PR 08-SEP-2000; 2000US-0231242.	
PR 08-SEP-2000; 2000US-0231243.	
PR 08-SEP-2000; 2000US-0231244.	
PR 08-SEP-2000; 2000US-0231413.	
PR 08-SEP-2000; 2000US-0231414.	
PR 08-SEP-2000; 2000US-0232080.	
PR 08-SEP-2000; 2000US-0232081.	
PR 12-SEP-2000; 2000US-0231968.	
PR 14-SEP-2000; 2000US-0232397.	
PR 14-SEP-2000; 2000US-0232398.	
PR 14-SEP-2000; 2000US-0232399.	
PR 14-SEP-2000; 2000US-0232401.	
PR 14-SEP-2000; 2000US-0233063.	
PR 14-SEP-2000; 2000US-0233064.	
PR 14-SEP-2000; 2000US-0233065.	
PR 21-SEP-2000; 2000US-0234223.	
PR 21-SEP-2000; 2000US-0234274.	
PR 25-SEP-2000; 2000US-0234997.	
PR 25-SEP-2000; 2000US-0234998.	
PR 26-SEP-2000; 2000US-0235484.	
PR 27-SEP-2000; 2000US-0235834.	
PR 27-SEP-2000; 2000US-0235835.	
PR 29-SEP-2000; 2000US-0236327.	
PR 29-SEP-2000; 2000US-0236367.	
PR 29-SEP-2000; 2000US-0236368.	
PR 29-SEP-2000; 2000US-0236369.	
PR 29-SEP-2000; 2000US-0236370.	
PR 02-OCT-2000; 2000US-0236802.	
PR 02-OCT-2000; 2000US-0237037.	
PR 02-OCT-2000; 2000US-0237038.	
PR 02-OCT-2000; 2000US-0237039.	
PR 02-OCT-2000; 2000US-0237040.	
PR 13-OCT-2000; 2000US-0239335.	
PR 13-OCT-2000; 2000US-0239337.	
PR 20-OCT-2000; 2000US-0240960.	
PR 20-OCT-2000; 2000US-0241221.	
PR 20-OCT-2000; 2000US-0241785.	
PR 20-OCT-2000; 2000US-0241786.	
PR 20-OCT-2000; 2000US-0241787.	
PR 20-OCT-2000; 2000US-0241808.	
PR 20-OCT-2000; 2000US-0241809.	
PR 20-OCT-2000; 2000US-0241826.	
PR 01-NOV-2000; 2000US-0244617.	
PR 08-NOV-2000; 2000US-0245474.	
PR 08-NOV-2000; 2000US-0246475.	
PR 08-NOV-2000; 2000US-0246476.	
PR 08-NOV-2000; 2000US-0246477.	
PR 08-NOV-2000; 2000US-0246478.	
PR 08-NOV-2000; 2000US-0246523.	
PR 08-NOV-2000; 2000US-0246524.	
PR 08-NOV-2000; 2	

KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX
OS Homo sapiens.
XX
PN WC200155314-A2.
FD 02-AUG-2000;
XX
PF 17-JAN-2001; 2001WO-US01324.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0190628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225269.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0225279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226684.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235936.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0237935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249257.
PR 17-NOV-2000; 2000US-0249259.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.

```
PR 11-DEC-2000; 2000US-0254397.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
FA (HUMA-): HUMAN GENOME SCI INC.
XX
FI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-502630/55.
XX
PT Polynucleotides encoding digestive system antigens, useful for
PT diagnosing, treating, preventing and/or prognosing disorders of the
PT digestive system, particularly cancer and cancer metastases -
XX
XX Disclosure; SEQ ID NO 2596; 386bp; English.
PS
XX
CC The present invention provides the protein and coding sequences of a
CC number of human digestive system antigens. These can be used in the
CC diagnosis, treatment and prevention of digestive system disorders,
CC including cancer, Meckel's diverticulum, bacterial or parasitic
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or
CC ulcerative colitis. The present sequence is a genomic DNA fragment
CC encoding a digestive system antigen of the invention.
XX
SQ Sequence 32152 BP; 9229 A; 7356 C; 7335 G; 8342 T; 0 other;
Query Match: 33.2%; Score 165.8; DP 22; Length 32132;
Best Local Similarity 74.6%; Pred. No. 2.5e-25;
Matches 244; Conservative 0; Mismatches 62; Indels 21; Gaps 2;
QY 192 TATTTATTTATTTATTTCTAGATGGAGCTCTTGGCTCTGTTGGCCAGCTGGATGCA 250
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
28891 TGTGTTTATTTTATTTTGTGAGACGGAGCTCTGGCTCTGTTGGCCAGCTGGATGCA 28812
QY 252 ATGGCATGATCTCGCTCACTGCAACCTGCACTTTCGGCTTCAAGCATCTCTCTGCT 311
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
28831 GTGGCATGATCTCGCTCACTGCAACCTGCGCTTCCAGCTTCATGGCTTCTCTGCT 28772
QY 312 CAATCTTCCAACTAGCTGGGACTACAGGATGAGCTGCGGACCTGGCGTGGTTGTTG 371
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
28771 CAGCTCCCGAGTAGCTGGGACTACAGGGGCTGCCACCACTCCGCTCAAAAGGTCG 28712
QY 372 TTTTAAA-----TTTGGAGGCGAGTGCAGTGGCCCATATCTGTG 412
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
28711 TTTTAAAGTACATATAAAATAATGTCTTGGGCGCAGGCGCAGTGGTCAACCTGTA 28652
QY 413 ATCCAGCACTTTGGGAGACCAAGGCGAGGCGATTACTTGGAGTCAAGGATTCAGAGCA 472
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
28651 ATCCAGCACTTTGGGAGGCGCAAGGTGGGTGAAT--CATGAGGTCAAGGATTCGAGAGCA 28594
QY 473 GCCAGGCCAATGGTAAACCATGTC 499
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
28593 ACCTGGCCACATGGTGAACCTGTC 28567
RESULT 15
AAK91534
ID AAK91534 standard; DNA; 32152 BP.
XX
AC AAK91534;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen genomic sequence SEQ ID NO: 5110.
XX
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ds.
XX
OS Homo sapiens.
XX
FN WO2001:55314-A2.
XX
PD 02-AUG-2001.
```

```
XX
PF 17-JAN-2001; 2001WO-US01324.
XX
PR 31-JAN-2003; 2003US-0179065.
PR 04-FEB-2003; 2003US-0180628.
PR 24-FEB-2003; 2003US-0184664.
PR 02-MAR-2003; 2003US-0186350.
PR 16-MAR-2003; 2003US-0189874.
PR 17-MAR-2003; 2003US-0190076.
PR 18-APR-2003; 2003US-0198123.
PR 19-MAY-2003; 2003US-0205515.
PR 27-JUN-2003; 2003US-0209467.
PR 28-JUN-2003; 2003US-0214886.
PR 30-JUN-2003; 2003US-0215135.
PR 27-JUL-2003; 2003US-0216647.
PR 27-JUL-2003; 2003US-0216880.
PR 11-JUL-2003; 2003US-0217487.
PR 11-JUL-2003; 2003US-0217496.
PR 14-JUL-2003; 2003US-0218290.
PR 26-JUL-2003; 2003US-0220963.
PR 26-JUL-2003; 2003US-0220964.
PR 14-AUG-2003; 2003US-0224518.
PR 14-AUG-2003; 2003US-0224519.
PR 14-AUG-2003; 2003US-0225213.
PR 14-AUG-2003; 2003US-0225214.
PR 14-AUG-2003; 2003US-0225266.
PR 14-AUG-2003; 2003US-0225267.
PR 14-AUG-2003; 2003US-0225268.
PR 14-AUG-2003; 2003US-0225270.
PR 14-AUG-2003; 2003US-0226279.
PR 22-AUG-2003; 2003US-0226681.
PR 22-AUG-2003; 2003US-0226868.
PR 22-AUG-2003; 2003US-0227182.
PR 23-AUG-2003; 2003US-0227009.
PR 30-AUG-2003; 2003US-0228924.
PR 01-SEP-2003; 2003US-0229287.
PR 01-SEP-2003; 2003US-0229343.
PR 01-SEP-2003; 2003US-0229344.
PR 01-SEP-2003; 2003US-0229345.
PR 03-SEP-2003; 2003US-0229509.
PR 03-SEP-2003; 2003US-0229513.
PR 06-SEP-2003; 2003US-0230437.
PR 06-SEP-2003; 2003US-0230438.
PR 08-SEP-2003; 2003US-0231242.
PR 08-SEP-2003; 2003US-0231243.
PR 08-SEP-2003; 2003US-0231244.
PR 08-SEP-2003; 2003US-0231413.
PR 08-SEP-2003; 2003US-0231414.
PR 08-SEP-2003; 2003US-0232080.
PR 08-SEP-2003; 2003US-0232081.
PR 12-SEP-2003; 2003US-0231968.
PR 14-SEP-2003; 2003US-0232397.
PR 14-SEP-2003; 2003US-0232398.
PR 14-SEP-2003; 2003US-0232399.
PR 14-SEP-2003; 2003US-0232400.
PR 14-SEP-2003; 2003US-0232401.
PR 14-SEP-2003; 2003US-0233063.
PR 14-SEP-2003; 2003US-0233064.
PR 14-SEP-2003; 2003US-0233065.
PR 21-SEP-2003; 2003US-0234223.
PR 21-SEP-2003; 2003US-0234274.
PR 25-SEP-2003; 2003US-0234997.
PR 25-SEP-2003; 2003US-0234998.
PR 26-SEP-2003; 2003US-0235484.
PR 27-SEP-2003; 2003US-0235834.
PR 27-SEP-2003; 2003US-0235836.
PR 29-SEP-2003; 2003US-0236327.
PR 29-SEP-2003; 2003US-0236367.
PR 29-SEP-2003; 2003US-0236368.
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 13:34:49 ; Search time 31.2916 seconds
(without alignments)
7052.739 Million cell updates/sec

Title: US-09-830-902-1_COPY_1_500
Perfect score: 500
Sequence: : taacagactctgtctgtgttt.....aacatggtaaaaccatgtcg 500

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 3
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 2%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq*
5: /cgn2_6/ptodata/2/ina/FOCUS_COMB.seq*
6: /cgn2_6/ptodata/2/ina/backfiles.seq*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
C 1	167.2	33.4	2480	4	US-09-534-638-3 Sequence 3, Appl
C 2	141.2	28.2	18000	4	US-09-657-346A-17 Sequence 17, Appl
C 3	136.4	27.3	45716	4	US-08-965-048-5 Sequence 5, Appl
C 4	136.4	27.3	45989	4	US-08-965-048-6 Sequence 6, Appl
C 5	136.4	27.3	47493	4	US-09-804-474A-3 Sequence 3, Appl
C 6	133.6	26.7	1268	4	US-09-369-247-42 Sequence 42, Appl
C 7	133.6	26.7	75395	4	US-09-984-890-3 Sequence 3, Appl
C 8	132.6	26.5	319608	4	US-09-539-332D-1 Sequence 1, Appl
C 9	132.6	26.5	319608	4	US-09-539-332D-1 Sequence 1, Appl
C 10	132.4	26.5	40000	4	US-09-780-049-18 Sequence 18, Appl
C 11	132	26.4	74962	4	US-09-685-853A-3 Sequence 3, Appl
C 12	132	26.4	90541	4	US-09-759-359A-3 Sequence 3, Appl
C 13	131.6	26.3	7130	3	US-09-056-105-31 Sequence 31, Appl
C 14	131.2	26.2	66804	4	US-09-74C-041-3 Sequence 3, Appl
C 15	131	26.2	1701	3	US-09-078-294-9 Sequence 9, Appl
C 16	130.8	26.2	55298	4	US-09-491-356C-1 Sequence 1, Appl
C 17	129.8	26.0	23187	4	US-09-499-522-1 Sequence 1, Appl
C 18	129.4	25.9	19650	4	US-09-819-989-3 Sequence 3, Appl
C 19	129.4	25.9	80246	3	US-09-078-294-4 Sequence 4, Appl
C 20	129.2	25.8	3885	1	US-08-688-145-1 Sequence 1, Appl
C 21	129.2	25.8	116592	4	US-09-818-512-3 Sequence 3, Appl
C 22	129	25.8	5375	3	US-08-757-223-7 Sequence 7, Appl
C 23	129	25.8	80595	3	US-09-078-294-3 Sequence 3, Appl
C 24	129	25.8	116592	4	US-09-818-512-3 Sequence 1, Appl
C 25	128.4	25.7	3001	4	US-09-539-332D-218 Sequence 218, App
C 26	128	25.6	863	3	US-08-943-731-171 Sequence 171, App
C 27	128	25.6	17606	3	US-08-943-731-4 Sequence 4, Appl

28	127.8	25.6	1764	4	US-09-620-312D-548 Sequence 548, App
29	127.8	25.6	4704	2	US-08-476-062A-52 Sequence 52, Appl
30	127.8	25.6	20201	4	US-09-734-674-3 Sequence 3, Appl
C 31	127.6	25.5	16063	4	US-09-801-052-3 Sequence 3, Appl
C 32	127	25.4	1460	4	US-09-904-615-44 Sequence 44, Appl
C 33	127	25.4	55298	4	US-09-491-356C-1 Sequence 1, Appl
C 34	126.8	25.4	62804	4	US-09-800-960-3 Sequence 3, Appl
C 35	126.4	25.3	896	3	US-08-943-731-31 Sequence 31, Appl
C 36	126.4	25.3	16609	3	US-08-943-731-1 Sequence 1, Appl
C 37	126.4	25.3	99500	4	US-09-796-036-10 Sequence 10, Appl
C 38	126.2	25.2	148567	4	US-09-801-875B-3 Sequence 3, Appl
C 39	126	25.2	50000	4	US-09-146-053-4 Sequence 4, Appl
C 40	126	25.2	319608	4	US-09-539-333F-1 Sequence 1, Appl
C 41	126	25.2	319608	4	US-09-679-439-1 Sequence 1, Appl
C 42	125.6	25.1	55827	4	US-09-813-133A-3 Sequence 3, Appl
C 43	125.4	25.1	1000	3	US-09-018-584A-33 Sequence 33, Appl
C 44	125.4	25.1	2099	3	US-08-938-669A-5 Sequence 5, Appl
C 45	125.4	25.1	2099	4	US-09-306-828-5 Sequence 5, Appl

ALIGNMENTS

RESULT :
US-09-534-638-3/c
: Sequence 3, Application US/09534638
: Patent No. 6320038
: GENERAL INFORMATION:
: APPLICANT: Panula, Periti A.J.
: APPLICANT: Brandt, Annika
: APPLICANT: Westerlund, Johanna
: TITLE OF INVENTION: Promoter for Neuropeptide FF Promoter and use thereof
: TITLE OF INVENTION: for therapy and diagnosis
: FILE REFERENCE: 2530-104
: CURRENT APPLICATION NUMBER: US/09/534,638
: EARLIER FILING DATE: 2000-03-27
: EARLIER APPLICATION NUMBER: 09/365755
: EARLIER FILING DATE: 1999-08-03
: NUMBER OF SEQ ID NOS: 22
: SOFTWARE: Patentin ver. 2.1
: SEQ ID NO 3
: LENGTH: 2480
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-534-638-3

Query Match	33.4%	Score	167.2	DB	4	Length	2480
Best Local Similarity	75.6%	Pred. No.	2.7e-307				
Matches	239	Conservative	0	Mismatches	63	Indels	14
Gaps	2						
QY	193	A	T	T	T	T	T
DB	1885	A	A	T	T	A	A
QY	252	A	T	G	C	A	T
DB	1825	G	T	G	G	C	G
QY	312	C	A	T	T	C	C
DB	1765	C	A	G	C	T	C
QY	372	T	T	T	T	A	A
DB	1705	T	T	T	T	A	A
QY	419	G	C	A	T	T	G
DB	1645	G	C	A	T	T	G
QY	479	C	C	A	C	A	T
DB	1595	C	C	A	C	A	T


```
RESULT 2
US-09-657-346A-17
; Sequence 17, Application US/09657346A
; Patent No. 6503754
; GENERAL INFORMATION:
; APPLICANT: Hong Zhang
; APPLICANT: Jacqueline Wyatt
; TITLE OF INVENTION: ANTISENSE MODULATION OF BDN INTERACTING DOMAIN DEATH AGONIST
; TITLE OF INVENTION: EXPRESSION
; FILE REFERENCE: RTS-0135
; CURRENT APPLICATION NUMBER: US/09/657,346A
; CURRENT FILING DATE: 2000-09-07
; NUMBER OF SEQ ID NOS: 174
; SEQ ID NO 17
; LENGTH: 18000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2144)...(2155)
; NAME/KEY: CDS
; LOCATION: (8247)...(8457)
; NAME/KEY: CDS
; LOCATION: (12772)...(12911)
; NAME/KEY: CDS
; LOCATION: (14031)...(14243)
; NAME/KEY: CDS
; LOCATION: (16669)...(16680)
US-09-657-346A-17
```

```
Query Match      28.2%; Score 141.2; DB 4; Length 18000;
Best Local Similarity 74.7%; Pred. No. 4.5e-24;
Matches 218; Conservative 0; Mismatches 68; Indels 6; Gaps 3;

QY 203 ATTTTATTTTCTAGATGGAGTCTCTCTGTGGCCAGGCTGGAAATGCAATGGCATGATC 262
Db 11355 ACTTTTAAAGAAACGGGGCTCTCTCTGTGGCCAGGCTGGAGTGTAGTGTATGATC 11414

QY 263 TCAGTCTACTGCAACTCCACCTTTCGGGTTCAGCAATCTCTGCTCCATCTTCCAA 322
Db 11415 ACAGTCTACTGCAAGCTTGAATCTGGCTTCAGCAATCTCTGCTCCATCTTCCAA 11474

QY 323 GTAGCTGGGACTACAGGAATGAGTGGCCAGCCTGGCTGGTGTGTGTAAATTT 382
Db 11475 GTAGCTGGGATTATAGGTACGTGC-GGTACACCTGGCTGGCTTTAAAGTGTTC-TGT 11530

QY 383 TGAGCCAGGTGCACTGCCCCATATCTGTATCCAGACACTTGGGAGAGCAAGGACGC 442
Db 11531 AGAGCAGGGCACAGTGGCTCACACTGTATCCAGACACTTGGGAGGCCCCAGGACGGA 11590

QY 443 CGATTACTTGAGGTGAGGAGTTCAAGACCCAGCCAGGCCACATCGGTAAAGCC 494
Db 11591 GGAACAC--AAGGTGAGGAGTTCGAGACCCAGCCCTGACCCACACATGGTGAAGCC 11640
```

```
RESULT 3
US-09-965-048-5
; Sequence 5, Application US/08965048
; Patent No. 6323244
; GENERAL INFORMATION:
; APPLICANT: Chen, Hong
; APPLICANT: Freimer, Nelson
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF NEUROPSYCHIATRIC DISORDERS
; FILE REFERENCE: 7853-093
; CURRENT APPLICATION NUMBER: US/08/965,048
; CURRENT FILING DATE: 1997-11-05
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 45716
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-965-048-5

Query Match      27.3%; Score 136.4; DB 4; Length 45716;
Best Local Similarity 71.2%; Pred. No. 7.2e-23;
Matches 195; Conservative 0; Mismatches 76; Indels 3; Gaps 1;

QY 106 CATTTTCACTGAAGTAACAATCTCTCCATAAAGTAGAGAAATCTCAAACTGGTGACTGGGA 165
Db 34614 CAATTGAGGTTAATAGACATGTTCTCTATGCTTGTAAATACAGAAATCTGACAGCAT 34673

QY 166 GTTTTGGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 222
Db 34674 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 34733

QY 223 TCTTGTCTCTGTTGGCCAGGCTGGAGTGAATGCAATGGCATGATCTCAGCTCACTGGCAACTCCA 282
Db 34734 TCTTACTCTGTGGCCAGGCTGGAGTGAATGCAATGGCATGATCTCGGCTCACTGGCAACTCTG 34793

QY 283 CCTTCGGGTTCAAGCAATCTCTCTGCTCCCTCAATCTTCCAACTAGCTGGGACTACAGGAAT 342
Db 34794 CCTCCGAGTTCAGCAATCTCTCTGCTCCCTCAGCTCCTGAGTAGCTGGGACTACAGATGT 34853

QY 343 GAGTGGCCGACCTGGGCTGGTTGTTGTTTGTGTTT 376
Db 34854 GTGCCACCATGCTGGCTTTTGTGTTTGTGTTT 34887
```

```
RESULT 4
US-09-965-048-6
; Sequence 6, Application US/08965048
; Patent No. 6323244
; GENERAL INFORMATION:
; APPLICANT: Chen, Hong
; APPLICANT: Freimer, Nelson
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF NEUROPSYCHIATRIC DISORDERS
; FILE REFERENCE: 7853-093
; CURRENT APPLICATION NUMBER: US/08/965,048
; CURRENT FILING DATE: 1997-11-05
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 45989
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-965-048-6
```

```
Query Match      27.3%; Score 136.4; DB 4; Length 45989;
Best Local Similarity 71.2%; Pred. No. 7.2e-23;
Matches 195; Conservative 0; Mismatches 76; Indels 3; Gaps 1;

QY 106 CATTTTCACTGAAGTAACAATCTCTCCATAAAGTAGAGAAATCTCAAACTGGTGACTGGGA 165
Db 34728 CAATTGAGGTTAATAGACATGTTCTCTATGCTTGTAAATACAGAAATCTGACAGCAT 34787

QY 166 GTTTTGGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 222
Db 34788 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 34847

QY 223 TCTTGTCTCTGTTGGCCAGGCTGGAGTGAATGCAATGGCATGATCTCAGCTCACTGGCAACTCCA 282
Db 34848 TCTTACTCTGTGGCCAGGCTGGAGTGAATGCAATGGCATGATCTCGGCTCACTGGCAACTCTG 34907

QY 283 CCTTCGGGTTCAAGCAATCTCTCTGCTCCCTCAATCTTCCAACTAGCTGGGACTACAGGAAT 342
Db 34908 CCTCCGAGTTCAGCAATCTCTCTGCTCCCTCAGCTCCTGAGTAGCTGGGACTACAGATGT 34967

QY 343 GAGTGGCCGACCTGGGCTGGTTGTTGTTTGTGTTT 376
Db 34968 GTGCCACCATGCTGGCTTTTGTGTTTGTGTTT 35001
```



```
1 NAME/KEY: exon
2 LOCATION: 216836..216994
3 OTHER INFORMATION: exon V
4 NAME/KEY: exon
5 LOCATION: 216836..217077
6 OTHER INFORMATION: exon V2
7 NAME/KEY: exon
8 LOCATION: 217671..217764
9 OTHER INFORMATION: exon V1
10 NAME/KEY: exon
11 LOCATION: 227655..227736
12 OTHER INFORMATION: exon V4
13 NAME/KEY: exon
14 LOCATION: 238715..238919
15 OTHER INFORMATION: exon V3
16 NAME/KEY: exon
17 LOCATION: 243440..243673
18 OTHER INFORMATION: exon W
19 NAME/KEY: exon
20 LOCATION: 240440..241153
21 OTHER INFORMATION: exon W2
22 NAME/KEY: exon
23 LOCATION: 241072..241291
24 OTHER INFORMATION: exon X
25 NAME/KEY: exon
26 LOCATION: 244353..244561
27 OTHER INFORMATION: exon Y
28 NAME/KEY: exon
29 LOCATION: 246273..247802
30 OTHER INFORMATION: exon Z
31 NAME/KEY: misc feature
32 LOCATION: 247803..249803
33 OTHER INFORMATION: 3'regulatory region
34 NAME/KEY: allele
35 LOCATION: 8316
36 OTHER INFORMATION: 99-27943-150 : polymorphic base G or C
37 NAME/KEY: allele
38 LOCATION: 21672
39 OTHER INFORMATION: 99-27935-193 : polymorphic base G or C
40 NAME/KEY: allele
41 LOCATION: 65485
42 OTHER INFORMATION: 8-128-33 : polymorphic base C or T
43 NAME/KEY: allele
44 LOCATION: 95396
45 OTHER INFORMATION: 99-31960-363 : polymorphic base A or G
46 NAME/KEY: allele
47 LOCATION: 107281
48 OTHER INFORMATION: 99-24656-260 : polymorphic base A or G
49 NAME/KEY: allele
50 LOCATION: 163640
51 OTHER INFORMATION: 99-24639-163 : polymorphic base A or C
52 NAME/KEY: allele
53 LOCATION: 163876
54 OTHER INFORMATION: 99-24634-108 : polymorphic base A or T
55 NAME/KEY: allele
56 LOCATION: 169974
57 OTHER INFORMATION: 99-7652-162 : polymorphic base A or G
58 NAME/KEY: allele
59 LOCATION: 170810
60 OTHER INFORMATION: 99-16100-147 : polymorphic base A or G
61 NAME/KEY: allele
62 LOCATION: 173358
63 OTHER INFORMATION: 99-5862-167 : polymorphic base A or G
64 NAME/KEY: allele
65 LOCATION: 189957
66 OTHER INFORMATION: 99-5919-215 : polymorphic base A or G
67 NAME/KEY: allele
68 LOCATION: 197163
69 OTHER INFORMATION: 99-24658-410 : polymorphic base A or G
70 NAME/KEY: allele
71 LOCATION: 200778
72 OTHER INFORMATION: 8-303-235 : polymorphic base A or G
73 NAME/KEY: allele
74 LOCATION: 202651
75 OTHER INFORMATION: 5-300-221 : polymorphic base A or G
76 NAME/KEY: allele
77 LOCATION: 202679
78 OTHER INFORMATION: 5-300-193 : polymorphic base A or G
79 NAME/KEY: allele
80 LOCATION: 203378
81 OTHER INFORMATION: 5-299-128 : polymorphic base A or T
82 NAME/KEY: allele
83 LOCATION: 204138
84 OTHER INFORMATION: 5-296-213 : polymorphic base A or T
85 NAME/KEY: allele
86 LOCATION: 204605
87 OTHER INFORMATION: 8-252-190 : polymorphic base C or T
88 NAME/KEY: allele
89 LOCATION: 204934
90 OTHER INFORMATION: 99-24644-194 : polymorphic base A or G
91 NAME/KEY: allele
92 LOCATION: 205206
93 OTHER INFORMATION: 8-295-248 : polymorphic base A or C
94 NAME/KEY: allele
95 LOCATION: 205329
96 OTHER INFORMATION: 8-295-125 : polymorphic base C or T
97 NAME/KEY: allele
98 LOCATION: 206064
99 OTHER INFORMATION: 8-293-130 : polymorphic base A or G
100 NAME/KEY: allele
101 LOCATION: 206545
102 OTHER INFORMATION: 8-292-198 : polymorphic base A or G
103 NAME/KEY: allele
104 LOCATION: 207313
105 OTHER INFORMATION: 8-251-322 : polymorphic base A or G
106 NAME/KEY: allele
107 LOCATION: 208285
108 OTHER INFORMATION: 8-289-322 : polymorphic base A or G
109 NAME/KEY: allele
110 LOCATION: 208960
111 OTHER INFORMATION: 5-287-249 : polymorphic base C or T
112 NAME/KEY: allele
113 LOCATION: 209123
114 OTHER INFORMATION: 8-287-86 : polymorphic base A or T
115 NAME/KEY: allele
116 LOCATION: 209631
117 OTHER INFORMATION: 8-285-319 : polymorphic base A or G
118 NAME/KEY: allele
119 LOCATION: 210361
120 OTHER INFORMATION: 8-283-278 : polymorphic base G or C
121 NAME/KEY: allele
122 LOCATION: 210463
123 OTHER INFORMATION: 8-283-176 : polymorphic base A or G
124 NAME/KEY: allele
125 LOCATION: 210486
126 OTHER INFORMATION: 8-283-153 : polymorphic base G or C
127 NAME/KEY: allele
128 LOCATION: 210583
129 OTHER INFORMATION: 8-283-56 : polymorphic base C or T
130 NAME/KEY: allele
131 LOCATION: 210879
132 OTHER INFORMATION: 8-282-345 : polymorphic base G or C
133 NAME/KEY: allele
134 LOCATION: 210964
135 OTHER INFORMATION: 8-282-260 : polymorphic base G or T
136 NAME/KEY: allele
137 LOCATION: 210979
138 OTHER INFORMATION: 8-282-245 : polymorphic base A or C
139 NAME/KEY: allele
140 LOCATION: 211050
141 OTHER INFORMATION: 8-282-174 : variable motif AAAGG or GAAGGAAGGAAGGAAGA
142 NAME/KEY: allele
143 LOCATION: 211132
144 OTHER INFORMATION: 8-282-92 : polymorphic base A or T
145 NAME/KEY: allele
146 LOCATION: 211247
```



```

1 Sequence 1611, Application US/16017161
2 Publication No. US20030143668A1
3 GENERAL INFORMATION:
4 APPLICANT: SUWA, MAKIKO
5 APPLICANT: ASAI, KIYOSHI
6 APPLICANT: AKIYAYA, YUTAKA
7 APPLICANT: ABURATANI, HIROYUKI
8 TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
9 FILE REFERENCE: 084335/0152
10 CURRENT APPLICATION NUMBER: US/10/017,161
11 PRIOR FILING DATE: 2002-12-18
12 PRIOR APPLICATION NUMBER: JP 2001/246789
13 NUMBER OF SEQ ID NOS: 2430
14 SOFTWARE: PatentIn Ver. 2.1
15 SEQ ID NO 1611
16 LENGTH: 39725
17 TYPE: DNA
18 ORGANISM: Homo sapiens
19 FEATURE:
20 NAME/KEY: source
21 LOCATION: (1)..(39725)
22 FEATURE:
23 NAME/KEY: CDS
24 LOCATION: (201)..(409)
25 FEATURE:
26 NAME/KEY: CDS
27 LOCATION: (1705)..(1966)
28 FEATURE:
29 NAME/KEY: CDS
30 LOCATION: (3861)..(4281)
31 FEATURE:
32 NAME/KEY: CDS
33 LOCATION: (6128)..(6513)
34 FEATURE:
35 NAME/KEY: CDS
36 LOCATION: (13450)..(13658)
37 FEATURE:
38 NAME/KEY: CDS
39 LOCATION: (15063)..(15253)
40 FEATURE:
41 NAME/KEY: CDS
42 LOCATION: (15939)..(16148)
43 FEATURE:
44 NAME/KEY: CDS
45 LOCATION: (20042)..(20167)
46 FEATURE:
47 NAME/KEY: CDS
48 LOCATION: (22121)..(22419)
49 FEATURE:
50 NAME/KEY: CDS
51 LOCATION: (22509)..(22680)
52 FEATURE:
53 NAME/KEY: CDS
54 LOCATION: (23434)..(23601)
55 FEATURE:
56 NAME/KEY: CDS
57 LOCATION: (24070)..(24306)
58 FEATURE:
59 NAME/KEY: CDS
60 LOCATION: (27349)..(27643)
61 FEATURE:
62 NAME/KEY: CDS
63 LOCATION: (28259)..(28404)
64 FEATURE:
65 NAME/KEY: CDS
66 LOCATION: (28906)..(29142)
67 FEATURE:
68 NAME/KEY: CDS
69 LOCATION: (30237)..(30376)
70 FEATURE:
71 NAME/KEY: CDS
72 LOCATION: (36366)..(36636)

```


Db 3322 GTGGCATGATCTCGGCTCACTGCAAGCTGGGCTTCCAGGTTTCATGCCATCTCTCTGCT 2381
Qy 312 CAATCTTCCAGTAGCTGGGACTACAGGAATAGGCTGCGGACCTGGCTGGTTGTTG 371
Db 3382 CAGCTCCCGAGTAGCTGGGACTACAGGCGCTGCCACACACCCGGTACAAAGGTGTA 3441
Qy 372 TTTTAA-----TTTGAAGGCCAGGTGCGAGTGGGCAATATCTGTG 412
Db 3442 TTTTAAAGTGACATAAAATAAATGCTTTGGGCGCAGGCGCAGTGGCTACACCTGTA 3501
Qy 413 ATCCAGCACTTTGGGAGACCAAGCCAGGCCGATTACTTGAAGTTCAGGAGTTCAAGACCA 472
Db 3502 ATCCAGCACTTTGGGAGGCCAAGGTGGTGAAT--CATGAGGTGAGGTTGAGAGCCA 3559
Qy 473 GCCAGGCCAACATGTTAAACCATGTC 499
Db 3560 ACCTGGCCAACATGTTGAACCCCTGTC 3586

RESULT 11
US-09-764-855-327
; Sequence 327, Application US/09764855
; Patent No. US20020119919A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA110
; CURRENT APPLICATION NUMBER: US/09/764,855
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 327
; LENGTH: 32204
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-855-327

Query Match 33.2%; Score 165.8; DB 10; Length 32204;
Best Local Similarity 74.6%; Pred. No. 4.6e-32;
Matches 244; Conservative 0; Mismatches 62; Indels 21; Gaps 2;
Qy 192 TATTTATTTTATTTTCTAGATGGAGTCTCTCTGTGGCCAGGCTGGAAATGCA 251
Db 8128 TGTITTTTTTTTTTTTTTTGAGACGGAGTCTCCCTCTGTGCCAGGCTGGAGTGCC 8187
Qy 252 ATGGCATGATCTCAGCTCACTGCAAGCTCCACCTTTCGGGTTCAGCAATCTCTGCT 311
Db 8189 GTGGCATGATCTCGGCTCACTGCAAGCTCGGCTTCCAGGTTTCATGCCATTTCTGCT 9247
Qy 312 CAATCTTCCAGTAGCTGGGACTACAGGAATAGGCTGCCGACCTGGCTGGTTGTTG 371
Db 8248 CAGCTCCCGAGTAGCTGGGACTACAGGCGCTGCCACACACCCGGCTACAAAGGTGTA 9307
Qy 372 TTTTAA-----TTTGAAGGCCAGGTGCGAGTGGGCAATATCTGTG 412
Db 8369 TTTTAAAGTGACATAAAATAAATGCTTTGGGCGCAGGCGCAGTGGCTACACCTGTA 8367
Qy 413 ATCCAGCACTTTGGGAGACCAAGCCAGGCCGATTACTTGAAGTTCAGGAGTTCAAGACCA 472
Db 8368 ATCCAGCACTTTGGGAGGCCAAGGTGGTGAAT--CATGAGGTGAGGTTGAGAGCCA 8426
Qy 473 GCCAGGCCAACATGTTAAACCATGTC 499
Db 8426 ACCTGGCCAACATGTTGAACCCCTGTC 8452

RESULT 12
US-09-764-872-517/c
; Sequence 510, Application US/09764872
; Publication No. US2003005023A1
; GENERAL INFORMATION:

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA125
; CURRENT APPLICATION NUMBER: US/09/764,872
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 957
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 517
; LENGTH: 32204
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-872-517
Query Match 33.2%; Score 165.8; DB 11; Length 32204;
Best Local Similarity 74.6%; Pred. No. 4.6e-32;
Matches 244; Conservative 0; Mismatches 62; Indels 21; Gaps 2;
Qy 192 TATTTATTTTATTTTCTAGATGGAGTCTCTCTGTGGCCAGGCTGGAAATGCA 251
Db 24077 TGTITTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTGGCCAGGCTGGAGTGCC 24015
Qy 252 ATGGCATGATCTCAGCTCACTGCAAGCTCCACCTTTCGGGTTCAGCAATCTCTGCT 311
Db 24017 GTGGCATGATCTCGGCTCACTGCAAGCTCGGCTTTCAGGTTTCATGCCATTTCTGCT 23958
Qy 312 CAATCTTCCAGTAGCTGGGACTACAGGAATAGGCTGCCGACCTGGCTGGTTGTTG 371
Db 23957 CAGCTCCCGAGTAGCTGGGACTACAGGCGCTGCCACACACCCGGCTACAAAGGTGTA 23898
Qy 372 TTTTAA-----TTTGAAGGCCAGGTGCGAGTGGGCAATATCTGTG 412
Db 23897 TTTTAAAGTGACATAAAATAAATGCTTTGGGCGCAGGCGCAGTGGCTACACCTGTA 23838
Qy 413 ATCCAGCACTTTGGGAGACCAAGCCAGGCCGATTACTTGAAGTTCAGGAGTTCAAGACCA 472
Db 23837 ATCCAGCACTTTGGGAGGCCAAGGTGGTGAAT--CATGAGGTGAGGTTGAGAGCCA 23780
Qy 473 GCCAGGCCAACATGTTAAACCATGTC 499
Db 23779 ACCTGGCCAACATGTTGAACCCCTGTC 23753

RESULT 13
US-10-072-349-327
; Sequence 327, Application US/10072349
; Publication No. US20030054420A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA110C1
; CURRENT APPLICATION NUMBER: US/10/072,349
; CURRENT FILING DATE: 2002-02-11
; Prior application removed - See file wrapper or Palm
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 327
; LENGTH: 32204
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-072-349-327

Query Match 33.2%; Score 165.8; DB 14; Length 32204;
Best Local Similarity 74.6%; Pred. No. 4.6e-32;
Matches 244; Conservative 0; Mismatches 62; Indels 21; Gaps 2;
Qy 192 TATTTATTTTATTTTCTAGATGGAGTCTCTCTGTGGCCAGGCTGGAAATGCA 251
Db 8128 TGTITTTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTGGCCAGGCTGGAGTGCC 8187
Qy 252 ATGGCATGATCTCAGCTCACTGCAAGCTCCACCTTTCGGGTTCAGCAATCTCTGCT 311
Db 8188 GTGGCATGATCTCGGCTCACTGCAAGCTCGGCTTCCAGGTTTCATGCCATTTCTGCT 9247

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Computer Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 18:12:56 : Search time 1044.61 Seconds
(without alignments)
11633.309 Million cell updates/sec

Title: US-09-830-902-1_copy_1_500
Perfect score: 500
Sequence: : taactgactctgtgtgtttt.....aacatggttaaacacatgtctg 510

Scoring table: IDENTITY_NJC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12:52238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:

1:	em_estha:
2:	em_esthuma:
3:	em_estin:
4:	em_estatu:
5:	em_estov:
6:	em_estpl:
7:	em_estro:
8:	em_htc:
9:	gb_est1:
10:	gb_est2:
11:	gb_htc:
12:	gb_est3:
13:	gb_est4:
14:	gb_est5:
15:	em_estfun:
16:	em_estom:
17:	em_gss_hum:
18:	em_gss_invi:
19:	em_gss_plni:
20:	em_gss_vrt:
21:	em_gss_fun:
22:	em_gss_mam:
23:	em_gss_mus:
24:	em_gss_pro:
25:	em_gss_fod:
26:	em_gss_phg:
27:	em_gss_vrl:
28:	gb_gss1:
29:	gb_gss2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	177.6	35.5	325	9 AI343123	AI343123 tb04e07.x
C 2	171.2	34.2	371	9 AW821209	AW821209 PM2-ST030
C 3	171.2	34.2	376	9 AW821105	AW821105 PM2-ST030
C 4	165.6	33.1	1027	10 BG622059	BG622059 602646665

5	165.4	33.1	439	12	BI467370
6	164.2	32.8	824	28	AQ888705
7	163.4	32.7	359	13	BQ353356
8	162.8	32.4	473	9	AA443390
9	157.8	31.6	681	29	AGC59079
10	152	30.4	445	9	AA573381
11	151.2	30.2	391	13	BQ574362
12	150.8	30.2	689	12	BM990649
13	148.8	29.8	347	14	CB266516
14	148.8	29.8	661	28	AQ539566
15	148.2	29.6	523	28	B69012
16	147.8	29.6	585	29	AG016069
17	146.6	29.3	416	13	BU685269
18	146.6	29.3	418	13	BU679484
19	146.6	29.3	1004	13	BX407955
20	146	29.2	418	10	BF931715
21	145.6	29.1	317	10	BE156651
22	145	29.0	367	12	BM682825
23	145	29.0	436	9	AA649183
24	144.6	28.9	492	28	B45413
25	144.2	28.8	627	29	AGC90691
26	143.8	28.8	479	9	AI863180
27	143.6	28.7	1052	13	BX463038
28	143	28.6	324	12	BM982363
29	142.8	28.6	353	9	AI815608
30	142.6	28.5	441	2	HSM085930
31	142.6	28.5	814	13	BU852129
32	142.4	28.5	502	2	HSM071632
33	142.2	28.4	475	9	AI261932
34	141.8	28.4	331	9	AI133357
35	141.8	28.4	639	9	AW163068
36	141.8	28.4	729	9	AV703572
37	141.6	28.3	711	29	AG011827
38	141.2	28.2	508	12	BM991594
39	141	28.2	787	29	BZ611164
40	140.6	28.1	360	9	AA847961
41	140.6	28.1	843	13	BU520971
42	140.4	28.1	404	14	CD251403
43	140.2	28.0	390	9	AI418927
44	139.8	28.0	827	12	BI457063
45	139.8	28.0	1379	11	BC022205

ALIGNMENTS

RESULT 1
AI343123/C
LOCUS
DEFINITION

AI343123 325 bp mRNA linear EST 16-FEB-1999
tb04e07.x2 NCI CGAP clone Homo sapiens cDNA clone IMAGE:2052612 3,
similar to contains Alu repetitive element; contains element PTRS
repetitive element ; mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AI343123
AI343123.1 GI:4060329
EST
Homo sapiens (human)

Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 325)

AUTHORS

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE
National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

cDNA Library Preparation: David B. Krizman, Ph.D.

cDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL

cDNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/bbrp/image/image.html

Insert Length: 437 Std Error: 0.00


```
VERSION BI467370.1 GI:15280248
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
1 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 439)
AUTHORS Melton, D., Brown, J., Kenny, G., Permutt, A., Lee, C., Kaestner, K.,
Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
Hillier, L., Marra, M., Pape, D., Wyllie, J., Martin, C., Blistain, A.,
Schmitt, A., Theising, B., Rutter, E., Ronko, J., Bennett, J., Cardenas
M., Gibbons, M., McCann, R., Cole, R., Tsagareishvili, R., Williams, T.,
Jackson, Y. and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished
Other_ESTs: ic23e02.y3
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biochem.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: Dr. Hiroshi Inoue
(hinoue@m.wustl.edu)
High quality sequence stop: 438.
FEATURES
Location/Qualifiers
1..439
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/tissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/clone_lib="HR85 islet"
/note="Organ: Pancreas; Vector: pBluescript SK(-); Site 1:
NotI; Site 2: XhoI; cDNA made by Cligo-dT priming.
Size-selected on agarose gel. Average insert size ~1kb. 5'
XhoI site was destroyed after directional cloning.
Amplified once. Contact information: Hiroshi Inoue, MD,
Metabolism Div. (Alan Permutt Lab), Washington University
School of Medicine, Box 8127, 660 South Euclid Ave., St.
Louis, MO 63110, E-mail: hinoue@gate.wustl.edu, Tel:
314-362-1916, Fax: 314-747-2692."
BASE COUNT 102 a 113 c 107 g 116 t 1 others
ORIGIN
Query Match 33.1%; Score 165.4; DB 12; Length 439;
Best Local Similarity 74.0%; Pred. No. 1.6e+03;
Matches 222; Conservative 0; Mismatches 77; Indels 1; Gaps 1;
QY 200 TTTATTTTATTTTATAGATGGAGTCTTCTGTTGCCACAGGTGGATGGATGGATG 259
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
76 TTTATTTTATTTTATAGACAGGGCTCACTCCGCTGCTCCAGGCTGGAGTGGCGA 135
QY 260 ATCTAGCTCACTGCAACCTCCACCTTTCCGGTTCAAGCAATCCCTCCCTCAATCTTC 319
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
136 GTACAGGTCAACCGCAACCTTGACCTGCTGGCTCAAGAAATGCTCCACCTCAGCCTCC 195
QY 320 CAAGTAGCTGGGACTACAGGAATGAGCTGCCGACCTGGCCCTGGTTGTTTATAA 379
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
196 CAAGTGTCTGGGACCACAGATGATGCCACCAATGTTGG-CTGATTTTACAAATTTTAA 254
QY 380 TTTTACAGCCAGCTGGCAGTGGCCCATATCTGTGATCCAGCACTTTGGGAGACCAAGCA 439
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
255 TTTTTCCTGGGCGCAGTGGCTCATGCTGTATCTCAGCACTTTGGGAGGCTGAGGCG 314
QY 440 GGCCGATTACTTGAGGTACAGGATTCAGACCAAGCCAGGCCAAGATGGTAAACCATGTC 499
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
315 GGCAGATCACCTGAGGTACAGGATTCAGACCAAGCCAGCTGGCCCAAGATGGCAAAACCTGTC 374
```

```
RESULT 6
AC888705 824 bp DNA linear GSS 10-NCV-1999
LOCUS HS3144 A2_E05_T70 CIT Approved Human Genomic Sperm Library 3 Homo
DEFINITION sapiens_genomic clone Plate=3144 Col=10 Row=I, genomic survey
sequence.
ACCESSION AC888705
VERSION AC888705.1 GI:63444895
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
1 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 824)
AUTHORS Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
Keller, A., Shaker, R., Furlong, C., Young, J., Zhao, S., Adams, M.D. and
Hood, L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE 99380589
PubMed 10449764
COMMENT Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3616
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end web server: http://www.htsc.washington.edu
Plate: 3144 row: I column: 10
Seq primer: T7
Class: BAC ends
High quality sequence stop: 824.
FEATURES
Location/Qualifiers
1..824
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone_plate="Plate=3144 Col=10 Row=I"
/sex="male"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/note="Organ: sperm; Vector: pReioBAC11; BAC clones in
E-Coli DH10B"
BASE COUNT 199 a 113 c 298 g 213 t 1 others
ORIGIN
Query Match 32.8%; Score 164.2; DB 28; Length 824;
Best Local Similarity 70.8%; Pred. No. 1.1e+03;
Matches 235; Conservative 0; Mismatches 89; Indels 8; Gaps 1;
QY 176 TGTGTTTGTGTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 235
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18 TGGTCTTTGAAGCAACTACTTAAATTTTATTTATTTTATTTATTTATTTATTTATTT 77
QY 236 CCAGGCTGGAATGCATGGCATGATCTCAGCTCACTGCAAGCTCCACCTTCGGGTTCA 295
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
78 TCCAGGCTGGCATGCAATGGCGAATCTCAGCTCACTGCAAGCTCTGCTCTCGGTTTA 127
QY 296 AGCAATCTCTCTGCTCAATCTTCCAAAGTAGTGGGACTACAGGAATGATGCGGCAAC 355
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
138 AGCAATCTCTCTGCTCAGCTCCCGAGTAGTGGGACTACAGSTGCCAGCCACTATGCC 197
QY 356 TGGCTGTTTGTGTTGT-----TTTAAATTTTAGGCCAGGTGCAGTGGCCCATAT 407
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
198 CGGTAATTTTGTATAATATATATATATATATATTTTGGTGGGCGACGGTGGCTCACAC 257
QY 408 CTGTGATCCAGCACTTTGGGAGACCAAGCGAGGCCGATTACTTTGAGGTCAAGGTTCAA 467
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
258 CTATAATCCAGCACTTTGGGAGGCTGAGGTGGGCGGATCACTTGAGGTCAAGGATGTAA 317
```


GenCore version 5.1.1f
Copyright (c) 1993 - 2003 CompuGen Ltd.

QM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:43:05 ; Search time 1384.09 Seconds
(without alignments)
14809.112 Million cell updates/sec

Title: US-09-830-902-1_copy_4000_4500

Perfect score: 501

Sequence: : ccaaaagtcgtggattacag.....tgaatgaggctagctaggcc 501

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045413386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:

1: gb_ba:

2: gb_htg:

3: gb_in:

4: gb_om:

5: gb_ov:

6: gb_pat:

7: gb_ph:

8: gb_pl:

9: gb_pri:

10: gb_rot:

11: gb_sts:

12: gb_sy:

13: gb_un:

14: gb_vit:

15: em_bat:

16: em_fun:

17: em_hum:

18: em_in:

19: em_mu:

20: em_or:

21: em_ov:

22: em_pat:

23: em_ph:

24: em_pi:

25: em_rc:

26: em_sts:

27: em_un:

28: em_vit:

29: em_htg_hum:

30: em_htg_inv:

31: em_htg_other:

32: em_htg_mus:

33: em_htg_pin:

34: em_htg_rtd:

35: em_htg_rtd:

36: em_htg_rtd:

37: em_htg_vrt:

38: em_sy:

39: em_htgo_hum:

40: em_htgo_mus:

41: em_htgo_other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	501	100.0	101584	9	CNS01DS5	AL121655 BAC seque
2	501	100.0	110000	6	AX093471	AX093471 Sequence
3	501	100.0	110000	9	HSA246003	AC246003 Homo sapi
4	501	100.0	155943	9	ACC12364	ACC12364 Homo sapi
5	98.8	19.7	127192	9	HS106502	AL035562 Human DNA
6	97.4	19.4	73096	2	AC091063	AC091063 Homo sapi
7	94	18.8	194835	9	AL158158	AL158158 Human DNA
8	93.6	18.7	142728	9	HSDJ792G4	ALC49636 Human DNA
9	85	17.0	75046	9	BX248112	BX248112 Human DNA
10	85	17.0	150770	9	AL953911	AL953911 Human DNA
11	85	17.0	187035	2	BX088588	BX088588 Homo sapi
12	84.4	16.8	63409	2	AC022992	AC022992 Homo sapi
13	84.4	16.8	97811	9	AC008634	AC008634 Homo sapi
14	84.4	16.8	104475	9	AC011394	AC011394 Homo sapi
15	84.2	16.8	121394	2	AL137799	AL137799 Homo sapi
16	84.2	16.8	190570	2	AL157894	AL157894 Homo sapi
17	83.8	16.7	93418	9	AC008379	AC008379 Homo sapi
18	83.8	16.7	179087	2	AC069489	AC069489 Homo sapi
19	83.8	16.7	198952	2	AP002016	AP002016 Homo sapi
20	83.4	16.6	61840	2	AC130291	AC130291 Homo sapi
21	83.4	16.6	132383	9	BX296545	BX296545 Human DNA
22	83.4	16.6	197676	2	AC022544	AC022544 Homo sapi
23	82.5	16.5	126613	9	HSJ603114	AL122001 Human DNA
24	82.2	16.4	163709	2	AP001155	AP001155 Homo sapi
25	82.2	16.4	185278	9	AC023043	AC023043 Homo sapi
26	82.2	16.4	195489	2	AP001159	AP001159 Homo sapi
27	82.2	16.4	210624	2	AP001321	AP001321 Homo sapi
28	81.8	16.3	84506	9	AL591667	AL591667 Human DNA
29	79.6	15.9	38476	9	U73024	U73024 Homo sapien
30	73.6	15.9	242825	9	HSPEX	Y10196 Homo sapien
31	73.2	15.9	137718	9	AL137066	AL137066 Human DNA
32	73.6	15.5	58562	2	AC104936	AC104936 Homo sapi
33	73.6	15.5	154893	2	AC072038	AC072038 Homo sapi
34	73.6	15.5	167106	9	CNS01RHA	AL161666 Human chr
35	73.6	15.5	193486	9	AC012123	AC012123 Homo sapi
36	73.4	15.4	89228	2	AC016273	AC016273 Homo sapi
37	73	15.4	163199	9	AC010096	AC010096 Homo sapi
38	72	15.4	178671	2	AC069283	AC069283 Homo sapi
39	76	15.2	101006	2	AC079392	AC079392 Homo sapi
40	76	15.2	11073	9	AC006375	AC006375 Homo sapi
41	75.8	15.1	100122	2	AF178220	AF178220 Homo sapi
42	75.8	15.1	123331	9	AC105416	AC105416 Homo sapi
43	75.8	15.1	155797	2	AC073049	AC073049 Homo sapi
44	75.8	15.1	157606	9	AC090897	AC090897 Homo sapi
45	75.8	15.1	194498	2	AC012359	AC012359 Homo sapi

ALIGNMENTS

RESULT 1
CNS01DS5
LOCUS
DEFINITION
BAC sequence from the SPG4 candidate region at 2p21-2p22 BAC 336P14
of library CtrB_978_SKB from chromosome 2 of Homo sapiens (Human).
ACCESSION
AL121655
VERSION
AL121655.1 GI:6002398
KEYWORDS
SPG4 genomic DNA interval.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 101584)
Hazan, G., Focknechten, N., Mavel, D., Paternotte, C., Sanson, D.,
Arriguenave, F., Davoine, C.S., Cruaud, C., Durr, A., Wincker, P.,

CNS01DS5 101584 bp DNA linear PRI 18-APR-2002
BAC sequence from the SPG4 candidate region at 2p21-2p22 BAC 336P14
of library CtrB_978_SKB from chromosome 2 of Homo sapiens (Human).
AL121655
AL121655.1 GI:6002398
SPG4 genomic DNA interval.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 101584)
Hazan, G., Focknechten, N., Mavel, D., Paternotte, C., Sanson, D.,
Arriguenave, F., Davoine, C.S., Cruaud, C., Durr, A., Wincker, P.,

QY 301 ACTTGCTTTGACAATGGGAATGTAGTACAAATGACACTGTGGCACTTTGGATTTAGGTT 360
|||||
Db 4300 ACTTGCTTTGACAATGGGAATGTAGTACAAATGACACTGTGGCACTTTGGATTTAGGTT 4360
|||||
QY 361 CGAGAGAACTTACACCTTCCACTCAGACTCTCTCTGGAAACCAAGATGCAATGAAGAAT 420
|||||
Db 4360 CGAGAGAACTTACACCTTCCACTCAGACTCTCTCTGGAAACCAAGATGCAATGAAGAAT 4410
|||||
QY 421 CAGGGTATCTCTGCTAGAGACATATGTCCAGCTTAATAGCCCAATCAACCTCTGACAT 480
|||||
Db 4420 CAGGGTATCTCTGCTAGAGACATATGTCCAGCTTAATAGCCCAATCAACCTCTGACAT 4470
|||||
QY 481 ATGAATGAGGCTAGCTAGGCC 501
|||||
Db 4480 ATGAATGAGGCTAGCTAGGCC 4500
|||||

RESULT 3
HSA246003 HSA246003 210000 bp DNA linear PRI 08-MAR-2000
LOCUS Homo sapiens Spast gene for spastin protein.
DEFINITION
ACCESSION AJ246003
VERSION AJ246003.1 GI:6273492
KEYWORDS Spast gene; spastin protein; SPG4-linked hereditary spastic paraplegia.
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Hazan, J., Fonknechten, N., Mavel, D., Paternotte, C., Samson, D., Artiguenave, P., Davoine, C., Cruaud, C., Durr, A., Winkler, P., Brottier, P., Cattolico, L., Barbe, V., Burgunder, J.M., Prud'homme, J.F., Brice, A., Fontaine, B., Heilig, R. and Weissenbach, J.
TITLE Spastin, a new AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia
JOURNAL Nat. Genet. 23 (3), 296-303 (1999)
MEDLINE 20055425
PUBMED 10610178
REFERENCE 2
AUTHORS Fonknechten, N., Mavel, D., Byrne, P., Davoine, C., Cruaud, C., Beentsch, D., Samson, D., Coutinho, P., Hutchinson, V., McMonagle, P., Burgunder, J., Tartaglione, A., Heinzel, C., Fekiz, Z., Deifeit, T., Parfrey, N., Brice, A., Fontaine, B., Prud'homme, J., Weissenbach, J., Durr, A. and Hazan, J.
TITLE Spectrum of SPG4 mutations in autosomal dominant spastic paraplegia
JOURNAL Hum. Mol. Genet. 9 (4), 637-644 (2000)
MEDLINE 20164302
PUBMED 10699187
REFERENCE 3 (bases : to 110000)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1999) Genoscope, Genoscope - Centre National de Sequencage, BP 191, EVRY 91006, FRANCE
COMMENT E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr. The sequence is the result of the assembly of 2 BAC clones: R-336p14 and 563N4, respectively from RPCI-11 and C12B_978_SKB library.
FEATURES
source
1..110000
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/map="2p21"
/clone="R-336P14"
/clone_lib="RPCI-11"
/note="BAC R-336P14 contains Spast gene: exon 1 to exon 3
BAC 563N4 contains Spast gene: exon 4 to exon 17"
9932..102009
/gene="Spast"

mRNA
9932..102009
/gene="Spast"
/product="spastin protein"
9932..10471
/gene="Spast"
/number=1
9932..10056
/gene="Spast"
9932..10056
/gene="Spast"
join(10057..10471,33719..33805,35749..35832,45023..45118,
60864..61051,61928..62061,62345..62438,73174..73248,
74634..74705,82789..82864,83103..83194,83335..83414,
88130..88172,89562..89641,91163..91233,93444..93484,
100600..100722)
/gene="Spast"
/cdon_start=1
/product="spastin protein"
/protein_id="CAB60208.1"
/db_xref="GI:6273493"
/translation="MNSPGSGKKGKGGGASNPVPPRPBPCLAPAPBAAGPAPBPES
PHRNLYFYSYDLFVGFALLRLVAFHJGLLFVWLCQRFRAKAAKRSAGAAPAPASA
SAPAPVPGGEAEKRVVFHKOAFEYISIALRIDEDKAGOEQAVWYKGIIELEKGI
AVITGQGEQCEERARRLOAKMTNLMVAKRLQLLEKMQPVLPFKSQDTDVYNDSTNL
ACRNGHLOSEGAVPKRKDPLHTNSLPRSKTVMKTSAGLSGHRRAFSYSGLSMVS
GVKQSGPAPTTHKGTPTKNTNKPSTPTTATRKKKOLKNFRNVDSNLINNEIVD
NGTAVKFDIAGQDLAKQALQIVILPSLRPELFTGLRAPARGLLJFGPPGNGKTMLA
KAVAAESNATFFNISAASTSKYVGEKLVRLFAVARELQPSIIFIDEVDSLLCER
REGHDASRLKTEFLIEFDGVQSAGDDRVLVMGATNRPOELDEAVLRRIKRVVSI
PNEETRLJLLKNLJCKQGSPLQKELAQALARMTCGYSGSULTALAKDALGPIRELKP
EQVKNMSASEMENRLSDFTESLKIKRSVSPQLEAYIRWNKDFGDTTV"
10472..33718
/gene="Spast"
/number=1
33719..33805
/gene="Spast"
/number=2
33806..35748
/gene="Spast"
/number=2
35749..35832
/gene="Spast"
/number=3
35833..45022
/gene="Spast"
/number=3
45023..45118
/gene="Spast"
/number=4
45119..60863
/gene="Spast"
/number=4
60864..61051
/gene="Spast"
/number=5
61052..61927
/gene="Spast"
/number=5
61928..62061
/gene="Spast"
/number=6
62062..62344
/gene="Spast"
/number=6
62345..62438
/gene="Spast"
/number=7
62439..73173
/gene="Spast"
/number=7
73174..73248
/gene="Spast"
/number=8
73249..74633
/gene="Spast"

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPi1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J.J. and de Jong, P.C. (1998). An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>), or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>).
VECTOR: pBACel.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-431P19; the clone sequenced to the right is RP11-78E13. Actual start of this clone is at base position 1 of RP11-444D15; actual end is at base position 155943 of RP11-444D15.

Data from AC011232 and AC010981 was used to finish this clone, AC012364. The sequence fidelity between bases 112626 to 112643 can not be guaranteed due to an unresolved homopolymeric run. The sequence between 112648 to 112663 is single stranded.

FEATURES

```
source
1..155943
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="2"
/maps="2"
/clone="RP11-444D15"
/clone_lib="RP11-11"
113..617
/rpt_family="Alu"
593..617
/rpt_family="AT_rich"
728..748
/rpt_family="AT_rich"
732..1026
/rpt_family="Alu"
1169..1240
/rpt_family="(ATATG)n"
1610..12503
/rpt_family="L2"
2504..12776
/rpt_family="Alu"
2777..13498
/rpt_family="L2"
3590..13742
/rpt_family="XIR"
4452..14765
/rpt_family="Alu"
4553..14975
/note="match to EST AA632695 (NID:G2556109) np83b02.s1"
4661..14692
misc_feature
```

```
repeat_region
4970..15219
/rpt_family="Alu"
5211..15218
/rpt_family="(CAAAA)n"
5234..15328
/note="match to EST BG571163 (NID:G13578829)"
5242..15328
/note="match to EST BG757662 (NID:G14068315)"
5247..15328
/note="match to EST AU121379 (NID:G10936614)"
5247..15328
/note="match to EST BF205430 (NID:G110399016)"
5247..15328
/note="match to EST BG033287 (NID:G12425431)"
5247..15328
/note="similar to Mus musculus EST AA619874 (NID:G2523750) v158932.s1"
5247..15328
/note="similar to Mus musculus EST AW120623 (NID:G6095956)"
5740..16346
/rpt_family="L2"
6347..16438
/rpt_family="AcHobo"
6453..16777
/rpt_family="J2"
6938..17406
/rpt_family="L1"
7415..18075
/rpt_family="ERV1"
7900..18421
/note="match to EST AW994939 (NID:G8255173)"
8087..18245
/rpt_family="L1"
8447..18524
/rpt_family="MER2_type"
8525..18799
/rpt_family="Alu"
8800..19067
/rpt_family="MER2_type"
9917..10069
/rpt_family="MER1_type"
10293..10786
/note="match to EST N54249 (NID:G1195415) yv65c12.s1"
10566..10753
/rpt_family="Alu"
10830..111973
/rpt_family="L1"
11589..111607
/rpt_family="(A)n"
11863..111888
/rpt_family="AT_rich"
11887..112044
/rpt_family="Tc2"
12045..112345
/rpt_family="Alu"
12382..112694
/rpt_family="Alu"
12656..112727
/rpt_family="A-rich"
12743..113296
/rpt_family="Tc2"
13463..113492
/rpt_family="(T)n"
13463..113478
/note="similar to Homo sapiens EST BG110835 (NID:G12604341)"
13820..114405
/note="match to EST BG434159 (NID:G133340653)"
14067..114097
/rpt_family="AT_rich"
14793..115032
repeat_region
```

```
repeat_region      /rpt_family="L1"
15333..15330
repeat_region      /rpt_family="Alu"
15331..15399
repeat_region      /rpt_family="L1"
15400..15700
misc_feature       /rpt_family="Alu"
15400..15410
repeat_region      /note="match to EST BG576844 (NID:gl3584497)"
15701..15745

Query Match:      100.0%; Score 501; DS 9; Length 155943;
Best Local Similarity 100.0%; Pred. No. 4,5e-113;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCAAGTGGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCGTTAAGCATGTTAATTAG 60
DB 119733 CCAAGTGGCTGGGATTACAGGCGTTAGCCACAGCATCCAGCGTTAAGCATGTTAATTAG 119792
QY 61 TTTTATTAATTCAGCAAAATGGTTGGAAATGCTGTCTTTAATGAGATGCTTAAGCTGCC 120
DB 119793 TTTTATTAATTCAGCAAAATGGTTGGAAATGCTGTCTTTAATGAGATGCTTAAGCTGCC 119832
QY 121 GTCTGAACATGAGGTAGAGGAATTTCTACACATTAATCATTTGCTGCTAATTAATCTTGCAA 180
DB 119853 GTCTGAACATGAGGTAGAGGAATTTCTACACATTAATCATTTGCTGCTAATTAATCTTGCAA 119912
QY 181 GATGGCCACACAAATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 240
DB 119913 GATGGCCACACAAATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 119972
QY 241 ACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTTGCACTAGAGTTGGCTTCTG 300
DB 119973 ACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTTGCACTAGAGTTGGCTTCTG 120032
QY 301 ACTTGTCTTGACAAATGGAATGTAGTACAAATGACACTGTGCACTTTGGATTTTAGGTTT 360
DB 120033 ACTTGTCTTGACAAATGGAATGTAGTACAAATGACACTGTGCACTTTGGATTTTAGGTTT 120092
QY 361 CGAGAGAACTTACACCTTCCACTCCACTCCACTCTCTCTGGAAACCAAGATGCAATGTAAGAAGT 420
DB 120093 CGAGAGAACTTACACCTTCCACTCCACTCCACTCTCTCTGGAAACCAAGATGCAATGTAAGAAGT 120152
QY 421 CAGGGCTATCTCTGTAGAGACATATGTCCAGCTAATAGCCCAATCAATCACTCTCTGAACAT 480
DB 120153 CAGGGCTATCTCTGTAGAGACATATGTCCAGCTAATAGCCCAATCAATCACTCTCTGAACAT 120212
QY 481 ATGAATGAGGCTAGTAGGCC 501
DB 120213 ATGAATGAGGCTAGTAGGCC 120233
```

```
RESULT 5
HS106502      127192 bp      DNA      linear      PRI 15-MAR-2001
LOCUS         Human DNA sequence from clone RPS-106502 on chromosome
DEFINITION    2p11.21-11.23 Contains a copy of the RPL41 gene for ribosomal
              protein L41, an ANT2 (fibroblast adenine nucleotide translocator 2
              (ADP-ATP carrier protein, ADP/ATP translocase)) pseudogene, the
              PAX1 gene for paired box gene 1, ESTs, STSs, GSSs, and three CpG
              islands, complete sequence.
ACCESSION     AL035562
VERSION       AL035562.14 GI:5123747
KEYWORDS      HTG; adenine nucleotide translocator; ANT2; CpG island; paired box;
              PAX1; ribosomal protein L41; RPL41.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 127192);
AUTHORS       Smith,M.
TITLE         Direct Submission
JOURNAL       Submitted (09-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire.
```

COMMENT

CS1015A, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jun 22, 1999 this sequence version replaced gi:4808216.
During sequence assembly data is compared from overlapping clones where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; information on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/Celegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
RPS-106502 is from the library RPL-5 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2

This sequence is the entire insert of clone RPS-106502.

FEATURES

source

Location/Qualifiers

1..127192
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="RZPD:RZP1P704O021065"
/db_xref="taxon:9606"
/chromosome="20"
/map="p11.21-11.23"
/clone="RPS-106502"
/clone_lib="RPL-5"
2..359
/note="match: GSS: Em:AQ224570"
1083..1406
/note="L1PB1 repeat: matches 5822..6152 of consensus"
complement(1407..1783)
/note="match: GSS: Em:AQ424123"
complement(1714..1746)
/note="match: GSS: Em:AQ284374"
2384..2931
/note="MLT1F repeat: matches 12..539 of consensus"
complement(3228..3649)
/note="match: GSS: Em:AQ176778"
3420..3642
/note="MER58A repeat: matches 1..224 of consensus"
3693..4352
/note="match: GSS: Em:AQ474310"
5113..5157
/note="L1ME3 repeat: matches 5746..5791 of consensus"
5158..5483
/note="ALUSC repeat: matches 1..304 of consensus"
5484..5868
/note="L1ME3 repeat: matches 5791..6163 of consensus"
6988..7110
/note="L1M4 repeat: matches 5694..5817 of consensus"
7135..7230
/note="MIR repeat: matches 113..211 of consensus"
7233..7520
/note="MLT1F repeat: matches 8..291 of consensus"
complement(7587..8085)
/note="match: GSS: Em:AQ205523"
8088..8376
/note="match: GSS: Em:AQ492709"
9472..9916

misc_feature

repeat_region

misc_feature

misc_feature

repeat_region

misc_feature

repeat_region

misc_feature

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

misc_feature

misc_feature

repeat_region

REFERENCE
AUTHORS

2 (bases 1 to 73096)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, M., Anderson, S.,
 Barna, N., Bastien, V., Boguski, L., Bouckgeater, B., Brown, A.,
 Camarata, C., Campopiano, A., Chang, J., Choquel, Y., Colangelo, M.,
 Collins, S., Collymore, A., Cooke, P., DeArrelano, K., Dewar, K.,
 Diaz, J. S., Dodge, S., Fero, S., Ferreira, P., FitzHugh, W., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Hagos, B., Heaford, A., Horton, E., Hulme, N.,
 Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRocque, K.,
 Lamarcares, R., Landers, T., Lechoczy, J., Levine, R., Liu, G.,
 MacLean, C., MacDonald, P., Marquis, N., Matthews, C., McCarthy, M.,
 McEwan, P., McKernan, K., McPheters, R., Medd, C., Meneses, S.,
 Mikova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C.,
 Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, G.,
 Peterson, K., Phunkhang, P., Fierre, N., Pollara, V., Raymond, C.,
 Retta, R., Rieback, M., Riley, R., Rice, C., Rogov, P., Roman, G.,
 Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S.,
 Severy, P., Sengere, C., Spencer, B., Stange-Thomann, N.,
 Stojanovic, N., Strauss, N., Subramanian, A., Talamas, C., Toffaye, S.,
 Theodore, J., Travers, M., Travis, N., Tzaghiolou, G., Vassiliou, H.,
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, K., Young, G.,
 Zainoun, C., Zembek, J., Zimmer, A. and Zedy, M.

TITLE

Submitted (25-MAR-2003) Whitehead Institute/MIT Center for Genome

JOURNAL

Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence@genome.wi.mit.edu

----- Project Information

Center project name: 102746

Center clone name: 1028_M_23

NOTE: This record contains 88 individual
 sequencing reads that have not been assembled into
 contigs. Runs of N are used to separate the reads
 and the order in which they appear is completely
 arbitrary. Low-pass sequence sampling is useful for
 identifying clones that may be gene-rich and allows
 overlap relationships among clones to be deduced.
 However, it should not be assumed that this clone
 will be sequenced to completion. In the event that
 the record is updated, the accession number will
 be preserved.

1 721: contig of 721 bp in length
 722 821: gap of 100 bp
 822 1559: contig of 738 bp in length
 1560 1659: gap of 100 bp
 1660 2369: contig of 720 bp in length
 2370 2469: gap of 100 bp
 2470 3207: contig of 738 bp in length
 3208 3307: gap of 100 bp
 3308 4097: contig of 790 bp in length
 4098 4197: gap of 100 bp
 4198 4925: contig of 728 bp in length
 4926 5025: gap of 100 bp
 5026 5765: contig of 740 bp in length
 5766 5865: gap of 100 bp
 5866 6409: contig of 344 bp in length
 6410 6509: gap of 100 bp
 6510 7252: contig of 743 bp in length
 7253 7352: gap of 100 bp
 7353 8096: contig of 744 bp in length
 8097 8196: gap of 100 bp
 8197 8935: contig of 739 bp in length
 8936 9036: gap of 100 bp
 9036 9784: contig of 749 bp in length
 9785 9884: gap of 100 bp
 9885 10623: contig of 739 bp in length

10624 10723: gap of 100 bp
 10724 11511: contig of 788 bp in length
 11512 11611: gap of 100 bp
 11612 12322: contig of 711 bp in length
 12323 12422: gap of 100 bp
 12423 13141: contig of 719 bp in length
 13142 13241: gap of 100 bp
 13242 13940: contig of 699 bp in length
 13941 14040: gap of 100 bp
 14041 14765: contig of 725 bp in length
 14766 14866: gap of 100 bp
 14866 15599: contig of 734 bp in length
 15600 15699: gap of 100 bp
 15699 16387: contig of 688 bp in length
 16388 16487: gap of 100 bp
 16487 17217: contig of 730 bp in length
 17218 17317: gap of 100 bp
 17318 18072: contig of 755 bp in length
 18073 18172: gap of 100 bp
 18173 18914: contig of 742 bp in length
 18915 19015: gap of 100 bp
 19015 19740: contig of 725 bp in length
 19740 20627: gap of 100 bp
 20628 20727: contig of 788 bp in length
 20728 21466: contig of 739 bp in length
 21467 21566: gap of 100 bp
 21567 22295: contig of 729 bp in length
 22296 23129: gap of 100 bp
 23129 23229: contig of 733 bp in length
 23229 23959: gap of 100 bp
 23960 24772: contig of 731 bp in length
 24773 24872: gap of 100 bp
 24873 25617: contig of 713 bp in length
 25618 25717: gap of 100 bp
 25718 26460: contig of 743 bp in length
 26461 26560: gap of 100 bp
 26561 27287: contig of 728 bp in length
 27287 28120: gap of 100 bp
 28121 28220: contig of 734 bp in length
 28221 28952: gap of 100 bp
 28953 29053: contig of 732 bp in length
 29053 29803: gap of 100 bp
 29803 30643: contig of 750 bp in length
 30644 30743: gap of 100 bp
 30744 31472: contig of 741 bp in length
 31472 31571: gap of 100 bp
 31572 32298: contig of 728 bp in length
 32299 32398: gap of 100 bp
 32399 33125: contig of 727 bp in length
 33126 33225: gap of 100 bp
 33226 33959: contig of 734 bp in length
 33960 34060: gap of 100 bp
 34060 34762: contig of 703 bp in length
 34763 34862: gap of 100 bp
 34863 35588: contig of 725 bp in length
 35588 36417: gap of 100 bp
 36417 37255: contig of 729 bp in length
 37256 37355: gap of 100 bp
 37356 38097: contig of 739 bp in length
 38097 38947: gap of 100 bp
 38948 39048: contig of 751 bp in length
 39048 39785: gap of 100 bp
 39786 40626: contig of 738 bp in length
 40627 40726: gap of 100 bp


```
repeat_region /note="VER13 repeat: matches 4. 175 of consensus"
4884. 5309
repeat_region /note="L2 repeat: matches 166. 616 of consensus"
5596. 5870
repeat_region /note="LIMA9 repeat: matches 6027. 6305 of consensus"
5871. 5938
repeat_region /note="Alu/FLAM repeat: matches 1. 68 of consensus"
5967. 6648
repeat_region /note="L2 repeat: matches 1284. 2090 of consensus"
6649. 6948
repeat_region /note="AluSq repeat: matches 1. 300 of consensus"
6949. 7003
repeat_region /note="L2 repeat: matches 2090. 2136 of consensus"
7072. 7462
repeat_region /note="L2 repeat: matches 2293. 2695 of consensus"
7588. 7789
repeat_region /note="MIR repeat: matches 12. 236 of consensus"
complement(5285. 8469)
misc_feature /note="match: STS: Em:G25863"
complement(8329. 8496)
misc_feature /note="match: GSS: Em:AQ227387"
complement(8329. 8495)
misc_feature /note="match: GSS: Em:B65752"
complement(8339. 8502)
misc_feature /note="match: GSS: Em:AQ110861"
6354. 8529
misc_feature /note="match: STS: Em:G54038"
Join(8370. 8532,9091. 9156,15133. 15192,15671. 15768)
repeat_region /note="match: STS: Em:G29920"
8591. 8662
repeat_region /note="36 copies 2 mer ta 81% conserved"
9732. 9893
repeat_region /note="AluJ repeat: matches 130. 287 of consensus"
complement(9737. 9894)
misc_feature /note="match: GSS: Em:AQ281101"
complement(9886. 10119)
misc_feature /note="match: GSS: Em:B53125"
complement(9897. 10091)
misc_feature /note="match: STS: Em:HS108WE1"
complement(9907. 10123)
misc_feature /note="match: GSS: Em:B17453"
9932. 9967
repeat_region /note="18 copies 2 mer ta 100% conserved"
9973. 10105
repeat_region /note="AluJ repeat: matches 1. 133 of consensus"
10133. 10392
repeat_region /note="AluSX repeat: matches 38. 299 of consensus"
10183. 10319
misc_feature /note="match: GSS: Em:AQ782105"
10404. 10492
repeat_region /note="MER5B repeat: matches 7. 97 of consensus"
10992. 11078
misc_feature /note="match: GSS: Em:AQ544271"
11354. 11653
repeat_region /note="AluSX repeat: matches 1. 300 of consensus"
11742. 12929
repeat_region /note="Tigger3b repeat: matches 2. 1231 of consensus"
12987. 13146
repeat_region /note="MIR repeat: matches 89. 248 of consensus"
13153. 13468
repeat_region /note="AluSX repeat: matches 1. 312 of consensus"
13678. 13848
repeat_region /note="L2 repeat: matches 2537. 2703 of consensus"
14394. 14639
repeat_region /note="MIR repeat: matches 12. 257 of consensus"
15816. 15902
misc_feature /note="match: GSS: Em:B46502"
15900. 15947
repeat_region /note="24 copies 2 mer aa 79% conserved"
16257. 16558
repeat_region /note="AluSq repeat: matches 1. 302 of consensus"
16956. 17220
repeat_region /note="AluSq repeat: matches 38. 304 of consensus"
```

```
repeat_region 17415. 17489
/note="37 copies 2 mer aa 71% conserved"
18552. 18845
repeat_region /note="AluSq repeat: matches 1. 293 of consensus"
19629. 19689
repeat_region /note="L2 repeat: matches 2685. 2743 of consensus"
19779. 21330
misc_feature /note="CpG island"
evidence=not_experimental
21481. 21679
repeat_region /note="MIR repeat: matches 48. 262 of consensus"
22591. 22818
repeat_region /note="MIR repeat: matches 23. 253 of consensus"
complement(23067. 23458)
misc_feature /note="match: GSS: Em:B43385"
complement(23074. 23456)
misc_feature /note="match: GSS: Em:B43132"
23731. 23893
repeat_region /note="VER5A repeat: matches 17. 182 of consensus"
24008. 24365
misc_feature /note="match: GSS: Em:AQ088039"
24107. 24196
repeat_region /note="MIR repeat: matches 105. 202 of consensus"
24953. 25091
repeat_region /note="LIMA8 repeat: matches 6145. 6287 of consensus"
25184. 25589
repeat_region /note="MER50 repeat: matches 1. 436 of consensus"
25620. 25718
repeat_region /note="MER50 repeat: matches 640. 734 of consensus"
25839. 25991
repeat_region /note="LIME1 repeat: matches 5212. 5366 of consensus"
25992. 26078
repeat_region /note="AluSp/Q repeat: matches 174. 260 of consensus"
26084. 26798
repeat_region /note="LIME1 repeat: matches 5347. 6036 of consensus"
27326. 27430
repeat_region /note="LIMB6 repeat: matches 4066. 4176 of consensus"
27431. 27740
repeat_region /note="AluSX repeat: matches 1. 310 of consensus"
27741. 28086
repeat_region /note="LIMB6 repeat: matches 4176. 4455 of consensus"
28087. 28226
repeat_region /note="AluJb repeat: matches 3. 132 of consensus"
28227. 28530
repeat_region /note="AluX repeat: matches 1. 297 of consensus"

Query Match 18.8% Score 94 DB 9 Length 194835
Best Local Similarity 62.4% Pred. No. 41e-13
Matches 181; Conservative 0; Mismatches 105; Indels 4; Gaps 2

QY 151 ACATAATCATTTGCTAAATTAATTCCTCAAGATGGCCACACAAATTCCTCTCTCAT 210
DB 33983 AAATAAGGACTAGAGTAGATTGCTTACAAAGATGGTCCCGATTCTCTAATATTAC 33924
QY 211 ATATATGCCCTTTGCAATGTGACTTTGCTACTTC-TCTATCAAGATGTGGAGTTATTT 269
DB 33923 AGGTCAGACCTTTGCAATATATACTGTGATACTTCTTGAATAAAGAGATGTAGTCTATT 33864
QY 270 TCCCATATATTGCACATAGATTGGCTTCTGACTTCTTGTGACATTTTGTGACAGATGAGTACAA 329
DB 33953 CCCCATCCCTTGAACCTGGATAGGCTTGTGACTTTTCTGACCAGTAGGTACAGCAGAA 33804
QY 330 ATGACACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAACTTACACCTTCCACTCACACT 389
DB 33803 GTAATATTGTAGGACTTACAAATCTAGACCCCAAGAAAGCTTGTAGCTTCCATTCTCACC 33744
QY 390 CTCTTGAACACCATGCAATGTAAAGAAAGTCAGGCTATCTCTCTAGAG 439
DB 33743 CTCTTCAACACAC---TGCCTTGAAGAAGCTCACTCTAACCCTACTGGAG 33657
```

RESULT 8
HSDJ792G4/c

LOCUS HSDJ792G4 142728 bp DNA linear PRI 31-DEC 2000
DEFINITION Human DNA sequence from clone RP4-792G4 on chromosome 1p31.2-32.2
Contains a transcription factor gene, a pseudogene similar to
tetraatricopeptide protein, ESTs, CA repeat (C15438), STSs, GSSs and
CpG islands, complete sequence.
ACCESSION AJ049636
VERSION AL049636.22 GI:11967799
KEYWORDS HTG; C15438.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 142728)
AUTHORS Howden, P.
TITLE Direct Submission
JOURNAL Submitted (20-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonereques@sanger.ac.uk
On Dec 22, 2000 this sequence version replaced GI:5531522.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EMBL, EMBL; SW, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/c_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr1>
RP4-792G4 is from the library RP4-4 constructed by the group of
Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP4-792G4.
FEATURES
source
location/Qualifiers
1..142728
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="p31.2-32.2"
/clone="RP4-792G4"
/clone_lib="RP4-4"
1..62
/note="Alu repeat: matches 1..62 of consensus"
1724..1797
/note="Charlie4 repeat: matches 1833..1856 of consensus"
1799..1842
/note="22 copies 2 mer at 84% conserved"
2387..2414
/note="14 copies 2 mer to 89% conserved"
10056..10503
/note="match: GSS: EM:AO165153"
13213..13497
/note="AluSg1 repeat: matches 1..309 of consensus"
13619..13930
/note="AluYa8 repeat: matches 1..308 of consensus"
14743..17426
/note="L1M1 repeat: matches -879..1556 of consensus"
17806..17891
/note="L1M1 repeat: matches 1540..1625 of consensus"
18385..19497
/note="L1M1 repeat: matches 1623..1761 of consensus"

misc_feature
23778..24163
/note="match: GSS: EM:AO270189"
24773..24820
/note="24 copies 2 mer at 81% conserved"
25736..25769
/note="17 copies 2 mer at 82% conserved"
27577..27630
/note="27 copies 2 mer at 74% conserved"
27634..27880
/note="MER2 repeat: matches 14..255 of consensus"
28476..28606
/note="AluSq/x repeat: matches 3..133 of consensus"
31013..31234
/note="L1MCI repeat: matches 5946..6172 of consensus"
31235..31323
/note="MER44C repeat: matches 1..89 of consensus"
31322..31503
/note="MER44C repeat: matches 235..417 of consensus"
31504..31670
/note="FRAX repeat: matches 1..168 of consensus"
31674..31962
/note="MER44C repeat: matches 413..697 of consensus"
31977..32132
/note="L1MCI repeat: matches 6169..6323 of consensus"
32644..32930
/note="AluYa5 repeat: matches 1..301 of consensus"
35861..35997
/note="AluSp/g repeat: matches 175..311 of consensus"
36422..36492
/note="L1T1-INTERNAL repeat: matches 808..881 of
consensus"
36772..36859
/note="L1T1-INTERNAL repeat: matches 1083..1157 of
consensus"
40721..40819
/note="MER33 repeat: matches 1..113 of consensus"
41156..41283
/note="MER33 repeat: matches 113..238 of consensus"
41326..41483
/note="L1MA2 repeat: matches 6151..6304 of consensus"
41795..41903
/note="L1MA2 repeat: matches 6241..6151 of consensus"
41904..41957
/note="YER33 repeat: matches 243..297 of consensus"
42270..42307
/note="19 copies 2 mer at 81% conserved"
44966..45522
/note="match: GSS: EM:AO422993"
46018..46404
/note="CpG island"
/evidence="not experimental"
complement(45601..47101)
/note="match: GSS: EM:AZ015581"
52353..53789
/gene="dJ792G4.1"
52353..53789
/gene="dJ792G4.1"
/note="match: CDNAS: EM:AF052249 EM:U39840 EM:J13220
EM:J59832 EM:AF071554 EM:U04198 EM:X01460 EM:AF052251
EM:U11690 EM:AF104902 EM:U37272 EM:J87333 EM:X92592
EM:X95603 EM:L38607 EM:L13202 EM:L13201 EM:X99291
EM:122760 EM:AB001572 EM:Y93128 EM:S79041 EM:AJ011652
EM:AF023915 EM:X74143 EM:U90538 EM:X55955 EM:S82462
EM:AF022869 EM:L12142 EM:L12143 EM:M14511 EM:J12703
EM:J13222 EM:AB028021
match: ESTs: EM:AI739311 EM:AI066718 EM:AI554075
EM:AA048965 EM:AI391745 EM:AI323408
match: proteins: Sw:P79772 Sw:Q61C60 Sw:Q16676 Sw:P11087
Sw:Q61345 Sw:Q64732 Sw:P55318 Sw:PC2452 Tr:Q93613
Sw:PC2453 Sw:Q01149 Tr:Q51415 Sw:P16791 Sw:P26372
Tr:Q92751 Tr:Q73784 Sw:Q99958 Sw:Q98937 Sw:Q12950
Sw:Q63244 Tr:Q70220 Sw:P06924 Sw:Q63245 Sw:Q64733
Tr:Q88778 Tr:Q73779 Wp:CE07013 Sw:P55787 Sw:Q60688


```

18187 AATTGAAGATGTCCTATACGGTCGATTTGGTTCCAAAGACGACTGCACACTGCATATCTTA 18246
204 TCCTCATATATATGCCCTTTGCAATGTGACCTTTGCACTTCTGCACTTCTGCACTGAGAGTGGAG 263
18247 TC-----TTACATGCCCTTTTSCAATATGACCTTTTSCACTTTTCCATTCACAGACATAGACT 18301
264 TTAATTTCCATATATTTGGCACTAGAGTTGGCCCTTCTGACTTGGCTTTGA-CAAATGGAATGT 322
18302 CTATCTTACCTCCCTTAAATCTGGACTAATCTTAATGAGTTGCTTTAACCAATATCATGT 18361
323 AGTACAAATGACACTGTGCAACTTTTGATTTTAGTTTGGAGAGAACTTACACCTTCCAC 382
18362 GGTAGAAATGACAAATGTGCAACTCCAAAGACAAAACTTCAAGAAATGCCACACACCTTCCAT 18421
383 TCACACTCTCTTTGGAACCAAGATGCAATG 411
18422 TTTTATGCTTTGGAATGCAGCTGCCATG 18450

RESULT 10
AL953911
LOCUS Human DNA sequence from clone RP11-242D12 on chromosome 9, complete
sequence.
AL953911
AL953911.4 GI:26780317
VERSION HTG.
KEYWORDS Homo sapiens (human);
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 150770)
Direct Submission
Submitted (13-DEC-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Dec 13, 2002 this sequence version replaced gi:25137359.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data files, paired quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest, except on the rare
occasion of the clone being a YAC.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
EMBL, EMBL; SWI, SWISSPROT; TrEMBL, TrEMBL; WPI, WORMPEP; information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormmap This sequence
was generated from part of bacterial clone contigs of human
chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr9
RP11-242D12 is from the library RPC1-11.1 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.
Location/Qualifiers

```

FEATURES

```

source 1. 150770
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-242D12"
/clone_db="RP11-11.1"
BASE COUNT 45356 a 30391 c 30483 g 44540 t
ORIGIN
Query Match 17.03; Score 85; DB 9; Length 150770;
Best local similarity 62.5%; Pred. NO. 7e-11;
Matches 168; Conservative 3; Mismatches 95; Indels 6; Gaps 2;
QY 144 ATTCTACACATATATCATTTGTGCTAAATTTACTTGCAAAGATGCCACACAAATTCCTCCTA 203
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
26 96018 AATTGAAGATGTCCTATACGGTCGATTTGTTCCAAAGACGACTGCACACTCCATCTCA 96077
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
204 TCCTCATATATATGCCCTTTGCAATGTGACCTTTGCACTTCTGCACTGAGAGTGGAGC 263
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
96078 TC-----TTACATGCCCTTTTSCAATATGACCTTTTCCACTTTTCCATTCACAGACATAGACT 96132
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
264 TTAATTTCCATATATTTGGCACTAGAGTTGGCCCTTCTGACTTGGCTTTGA-CAAATGGAATGT 322
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
96133 CTATCTTACCTCCCTTAAATCTGGACTAATCTTAATGAGTTGCTTTAACCAATATCATGT 96192
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
323 AGTACAAATGACACTGTGCAACTTTTGATTTTAGTTTGGAGAGAACTTACACCTTCCAC 382
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
96193 GGTAGAAATGACAAATGTGCAACTCCAAAGACAAAACTTCAAGAAATGCCACACCTTCCAT 96252
QY ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
96253 TTTTATGCTTTGGAATGCAGCTGCCATG 96281

RESULT 11
BX088588/c
LOCUS BX088588 187035 bp DNA linear HTG 08-APR-2003
DEFINITION Homo sapiens chromosome 9 clone RP11-5019, *** SEQUENCING IN
PROGRESS ***.
ACCESSION BX088588
VERSION BX088588.4 GI:27805046
KEYWORDS HTG; HTGS_PHASE2.
SOURCE Homo sapiens (human);
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 187035)
Plumb.B.
Direct Submission
Submitted (07-APR-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jan 20, 2003 this sequence version replaced gi:27803285.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project information
Center project name: BA5013
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 187023 bases at least Q40
Consensus quality: 187031 bases at least Q30
Consensus quality: 187034 bases at least Q20
Insert size: 187035; sum-of-contigs
Insert size: 196616; 2.8% error; agarose-fp
Quality coverage: 16.59x in Q20 bases; sum-of-contigs Quality
coverage: 15.87x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently

```

- * consists of 1 contigs. Gaps between the contigs
- * are represented as runs of N. The order of the pieces
- * is believed to be correct as given, however the sizes
- * of the gaps between them are based on estimates that have
- * provided by the submitter.
- * This sequence will be replaced
- * by the finished sequence as soon as it is available and
- * the accession number will be preserved.
- * 187035: contig of 187035 bp in length.

Location/Qualifiers

FEATURES

source

1..187035
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/clone="RP11-5019"
/clone_lib="RPC1-11.1"

misc_feature

1..187035

/note="assembly fragment:01460"

BASE COUNT 55353 a 37202 c 37514 g 56966 t

ORIGIN

Query March 17.0%; Score 85; DB 2; Length 187035;
Best Local Similarity 62.5%; Pred. No. 6.8e-12;
Matches 168; Conservative 0; Mismatches 95; Indels 6; Gaps 2;
QY 144 ATCTACACATAATCATTTGCTAAATTACTTCCAAAGATGCGACACACATTCCTCCN 203
DQ 125154 AATTGAAGATGTCCTATACGGTGCATTGTTCCAAAGACGACTGCACACTCCATTCTA 125235
QY 204 TCCATATATATGCGCCCTTTGCAATGTGACTTGTCTTCTATCATCATGATGGAGC 243
DQ 125294 TC-----TTACATGCGCCCTTTGCAATATGACTTTCCACTTTCCATCCACACATGACT 125340
QY 264 TTATTTCCCATATATTCGACTAGACTTGGCGCTTCTGACTTGGTTTGA CAATGGAATGT 322
DQ 125239 CTATCTTACCTCCCTTAAATCTGGACTAATCTCATGATGATGCTTTAACCAATATCATGT 125180
QY 323 AGTACAAATGACACTGTGCAACTTTTGATTTTGGTTTCGAGAGAACTTACACCTTCCAC 382
DQ 125179 GGTAGAATGACAAATGTGCAACTTCCAAAGACAAAACTTCAAGAAATCCCAACCTTCCAT 125120
QY 383 TCACACTCTCTTGGAAACCAAGATGCAATG 411
DQ 125119 TTTTATGCTTTGGAATGCAGCTGCCATG 125091

RESULT 12

AC022982/c

LOCUS

DEFINITION

Homo sapiens chromosome 18 clone RP11-545P23 map 18, LOW-PASS

SEQUENCE SAMPLING.

AC022982

AC022982.2 GI:9161727

KEYWORDS

HTG; HTGS_PHASE0.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 63409)

Unpublished

2 (bases 1 to 63409)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,

Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,

Chapel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,

DeArelanc, K., Dewar, K., Domino, M., Doyle, M., Fenestor, J.,

Ferreira, P., FitzHugh, W., Forrest, C., Gage, S., Galagan, J.,

Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,

Howland, J.C., Johnson, R., Jones, C., Kahn, L., Karas, A., Klein, J.,

Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, S., Locke, K.,

Macdonald, P., Margulis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheeters, R., Meddrim, C., Meneus, J., Morrow, J., Naylor, J.,
Norman, C.H., O'Connor, T., O'Donnell, P., Olivari, T.M., Peterson, K.,
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, A., Tesfaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.,
Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

TITLE Direct Submission
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 97811)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (23-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Oct 23, 2001 this sequence version replaced gi:19256017.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence
Estimated Total Number of Errors is 0.
Features
Location/Qualifiers
1..97811
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clones="CTB-174K1.4"
BASE COUNT 30500 a 19379 c 18444 g 29488 t
ORIGIN
Query Match 16.8%; Score 84.4; DB 9; Length 97811;
Best Local Similarity 52.6%; Pred. No. 1.1e-10;
Matches 164; Conservative 0; Mismatches 96; Indels 2; Gaps 2;
QY 138 AAGGAAATTCACACATAATCATGTGCTAAATTACTTGCAGAGATGGCCAGACAAATTC 147
DB 87692 AAGTCTATTATATTACTCAGATGGAAGGTGTGTTACCAAGGTGGCCAGACAAATTA 87952
QY 198 CTCCTATCCTCATATATATGCCCCCTTGCAGATGTGACTTGTCTCTCTATCATAGATG 257
DB 57752 TTCCCATCCCTATATCCAGGACGCTTGGCTGTGTGACTTCACCACTCCTCTCAGAG-GA 87810
QY 259 TGGAGCTTATTTTCCCATATATTGCACATAGAGTTGGCCTTGTGACTTGTCTTGA-CAATG 316
DB 87811 AAGAGTCTATTTTACTTCCCTTGAATTAAACTGTCTCTGTGATTTGCTTGACCAATA 87870
QY 317 GAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGTTTCGAGAGAACTTACACC 376
DB 87871 GAAGGTGTACAGATGACACAGGACAACTTCTGGGCTAGGTCTCAGAGGCGATCTCAAC 87930
QY 377 TTCACACTCACACTCTCTTTGGAA 398
DB 87931 ATCTGTTTTCAGCCCAATGGAA 87952
RESULT 14
AC011384/c
LOCUS AC011384 104475 bp DNA linear PRI 07-DEC-2001
DEFINITION Homo sapiens chromosome 5 clone CTB-143D11, complete sequence.
ACCESSION AC011384
VERSION AC011384.4 GI:17402759
KEYWORDS HTG.
SOURCE Homo sapiens (human);
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 104475)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 104475)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (06-OCT-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 104475)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission

JOURNAL Submitted (07-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Dec 7, 2001 this sequence version replaced gi:13752665.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence
Estimated Total Number of Errors is 0.
Features
Location/Qualifiers
1..104475
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTB-143D11"
BASE COUNT 31967 a 19128 c 19721 g 33609 t
ORIGIN
Query Match 16.8%; Score 84.4; DB 9; Length 104475;
Best Local Similarity 52.6%; Pred. No. 1.1e-10;
Matches 164; Conservative 0; Mismatches 96; Indels 2; Gaps 2;
QY 138 AAGGAAATTCACACATAATCATGTGCTAAATTACTTGCAGAGATGGCCAGACAAATTC 197
DB 94402 AAGTCTATTATATTACTCAGATGGAAGGTGTGTTACCAAGGTGGCCAGACAAATTA 94343
QY 198 CTCCTATCCTCATATATATGCCCCCTTGCAGATGTGACTTGTCTCTCTATCATAGATG 257
DB 94342 TTCCCATCCCTATATCCAGGACGCTTGGCTGTGTGACTTCACCACTCCTCTCAGAG-GA 94284
QY 258 TGGAGCTTATTTTCCCATATATTGCACATAGAGTTGGCCTTGTGACTTGTCTTGA-CAATG 316
DB 94283 AAGAGTCTATTTTACTTCCCTTGAATTAAACTGTCTCTGTGATTTGCTTTGACCAATA 94224
QY 317 GAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGTTTCGAGAGAACTTACACC 376
DB 94223 GAAGGTGTACAGATGACACAGGACAACTTCTGGGCTAGGTCTCAGAGGCGATCTCAAC 94164
QY 377 TTCACACTCACACTCTCTTTGGAA 398
DB 94163 ATCTGTTTTCAGCCCAATGGAA 94142
RESULT 15
AL137799
LOCUS AL137799 121394 bp DNA linear HTG 10-JUL-2001
DEFINITION Homo sapiens chromosome 1 clone RP5-1199A9, *** SEQUENCING IN PROGRESS ***, 6 unordered pieces.
AL137799
AL137799.4 GI:10039491
VERSION AL137799.4
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Plumb,B.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Sep 8, 2003 this sequence version replaced gi:19211917.
COMMENT
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: dJ1199A8
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; LC8752; 100% of reads

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Coriagen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:39:20 ; Search time 126.251 Seconds
(without alignments:
10712.169 Million cell updates/sec

Title: US-09-830-902-1_COPY_4000_4500

Perfect score: 501

Sequence: 1 ccaaaagtgcggatcacag.....tgatgaggtagtaggc 501

Scoring table: IDENTITY NJC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2004.DAT.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	501	100.0	110000	22	AAF84800
C 2	82.8	16.5	502	20	AAZ22641
C 3	82.6	16.5	854	20	AAZ22642
C 4	82.6	16.5	27869	22	ABA19635
C 5	82.6	16.5	27869	22	AAK66517
C 6	70	14.0	218336	25	ABQ76678
C 7	64.6	12.9	5239	24	ABL32273
C 8	63	12.6	1241	21	AAC74337

9	41.4	12.3	1473	27	AAC05322	Human secreted pro
10	41.4	12.3	1484	22	AAC05372	Human secreted pro
11	62	12.2	670	22	AAS46311	Human G protein-co
12	62	12.2	670	24	ABK81689	cDNA encoding nove
13	60.4	12.1	5239	24	ABL32272	Human immune syste
C 14	60.2	12.0	652	20	AAV88847	EST clone G151. H
C 15	59.4	11.5	2477	22	AAH17355	Human cDNA sequenc
16	56.8	11.3	1040	21	AAA81636	N. meningitidis pa
17	56.6	11.3	567571	25	AAS33224	Human chromosome 3
18	56.4	11.3	1737	23	ABX12157	cDNA encoding huma
C 19	56.4	11.3	5434	22	AAK76210	Human immune/haema
20	53.8	10.7	3535	22	AAK65528	Human immune/haema
21	53.8	10.7	3535	22	AAK73493	Human immune/haema
22	53.6	10.7	1214	22	AAI59283	Human polynucleoti
23	53.6	10.7	1275	22	AAI61369	Human polynucleoti
24	53.6	10.7	1586	21	AAZ86793	Human protein kina
25	53.6	10.7	1586	22	AAD11844	Human protein kina
26	53.6	10.7	1586	24	ABQ76287	Human protein kina
27	53.6	10.7	1868	24	AAS17048	Human protein kina
28	53.6	10.7	1868	25	ABX14972	Human protein kina
29	53.6	10.7	2142	22	AAH42064	Cell cycle protein
C 30	53.6	10.7	2375	22	AAS01478	Human secreted pro
31	53.6	10.7	2538	22	AAH76219	Human kinase PKIN-
32	53.6	10.7	3240	22	AAF75345	Human TGF-beta rec
33	53.6	10.7	3387	22	AAF44687	Novel protein kina
34	53.2	10.6	90220	24	ABK83576	Human cDNA differe
C 35	53.2	10.6	100000	24	ABQ74541	Human transglutari
C 36	52.6	10.5	397658	25	ABQ83210	Human transporter
37	52.2	10.4	2141	22	AAH42063	Cell cycle protein
38	51.6	10.3	245	21	AAC31581	Human secreted pro
39	51.6	10.3	1434	22	AAK89069	Human digestive sy
40	51.6	10.3	1434	22	AAI57626	Human colorectal c
41	51.6	10.3	1434	24	ABS99803	Genomic DNA #7 enc
42	51	10.2	502	23	ABV05821	Human prostate exp
C 43	50.4	10.1	10137	22	ABA18984	Human nervous syst
C 44	50.2	10.0	389	22	AAK64465	Human immune/haema
C 45	50.2	10.0	972	22	AAK73675	Human immune/haema

ALIGNMENTS

RESULT 1		AAF84800		AAF84800 standard; DNA: 110000 BP.	
10	AAF84800	standard; DNA: 110000 BP.			
XX	AAF84800				
AC	AAF84800				
XX	00-JUL-2002 (first entry)				
XX	Nucleotide sequence of the human SPG4 gene.				
XX	Human: SPG4 gene; spastin; PSF-AD; gene therapy; autosomal dominant familial spastic paraplegia; ss.				
XX	Homo sapiens.				
XX					
XX	Key	Location/Qualifiers			
FT	CDS	3932..102003			
FT		/*tag= a			
FT	exon	/note= "contains introns"			
FT		9932..10471			
FT		/*tag= b			
FT		/number= "1"			
FT	intron	10472..33718			
FT		/*tag= c			
FT		/number= "1"			
FT	exon	33719..33805			
FT		/*tag= d			
FT		/number= "2"			
FT	intron	33806..35748			
FT		/*tag= e			
FT		/number= "2"			

FT 35749..35832
FT /tag= f
FT /number= "3"
FT 35833..45022
FT /tag= g
FT /number= "3"
FT 45023..45118
FT /tag= h
FT /number= "4"
FT 45119..60863
FT /tag= i
FT /number= "4"
FT 60864..61051
FT /tag= j
FT /number= "5"
FT 61052..61927
FT /tag= k
FT /number= "5"
FT 61928..62061
FT /tag= l
FT /number= "6"
FT 62062..62344
FT /tag= m
FT /number= "6"
FT 62345..62438
FT /tag= n
FT /number= "7"
FT 62439..73173
FT /tag= o
FT /number= "7"
FT 73174..73248
FT /tag= p
FT /number= "8"
FT 73249..74633
FT /tag= q
FT /number= "8"
FT 74634..74705
FT /tag= r
FT /number= "9"
FT 74706..82788
FT /tag= s
FT /number= "9"
FT 82789..82864
FT /tag= t
FT /number= "10"
FT 82865..83102
FT /tag= u
FT /number= "10"
FT 83103..83194
FT /tag= v
FT /number= "11"
FT 83195..83334
FT /tag= w
FT /number= "11"
FT 83335..83414
FT /tag= x
FT /number= "12"
FT 83415..88129
FT /tag= y
FT /number= "12"
FT 88130..88172
FT /tag= z
FT /number= "13"
FT 88173..89561
FT /tag= aa
FT /number= "13"
FT 89562..89641
FT /tag= ab
FT /number= "14"
FT 89642..91162
FT /tag= ac
FT /number= "14"
FT 91163..91233

FT /tag= ad
FT /number= "15"
FT 91234..93443
FT /tag= ae
FT /number= "15"
FT 93444..93484
FT /tag= af
FT /number= "16"
FT 93485..100599
FT /tag= ag
FT /number= "16"
FT 100600..102009
FT /tag= ah
FT /number= "17"

XX FR2799138-A1.
XX 09-XAF-2001.
XX 03-SEP-1999; 99FR-0011097.
XX 03-SEP-1999; 99FR-0011097.
XX (CNRS : CNRS CENT NAT RECH SCI.
XX Weissenbach J, Hazan J;
XX WPI: 2001-283966730.
XX New human nucleic acid from the SPG4 gene, useful e.g. for diagnosis of
XX autosomal dominant familial spastic paraplegia and in drug screening -
XX Clair 2; Page 45-106; 145pp; French.
XX The present sequence represents a human SPG4 gene. The SPG4 gene encodes
XX a spastin polypeptide. Mutations in the SPG4 gene are responsible for
XX autosomal dominant familial spastic paraplegia. SPG4 polynucleotides,
XX and their fragments, are used to screen DNA banks for sequences that
XX encode spastin (particularly sequences in other mammals, specifically
XX mice); to identify SPG4 mutations, or other genetic anomalies,
XX particularly for diagnosis of autosomal dominant familial spastic
XX paraplegia (PSP-AD); to identify promoters and other regulatory elements
XX of the SPG4 gene; for detection and amplification; for recombinant
XX production of spastin; and for diagnostic genotyping of PSP-AD.
XX Sequence 1:10000 BP; 30622 A; 21640 C; 22917 G; 34921 T; 0 other;

Query Match 100.0%; Score 501; DB 22; Length 110000;
Best Local Similarity 100.0%; Pred. No. 8,1e-137;
Matches 501; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 CCAAAGTGTGGATTACAGGGCTTAGCCACAGCATCCAGCCTTAAGCATGTTAAATTAAAG 60
DB 4000 CCAAAGTGTGGATTACAGGGCTTAGCCACAGCATCCAGCCTTAAGCATGTTAAATTAAAG 4059
QY 61 TTTTATTAATTCAGCAAAATGGTTGGAAAATGCTGCTTAAATGAGATGCTTAAGCTGCC 120
DB 4050 TTTTATTAATTCAGCAAAATGGTTGGAAAATGCTGCTTAAATGAGATGCTTAAGCTGCC 4119
QY 121 GTCTGACATGAGGTAGAGGAAATTCACACATATATATGCTTAAATTAATTACTTGCAAA 180
DB 4120 GTCTGACATGAGGTAGAGGAAATTCACACATATATATGCTTAAATTAATTACTTGCAAA 4179
QY 181 GATGCCCAACAATTCCTCTATCTCATATATATATGCTTAAATTAATTACTTGCAAA 240
DB 4180 GATGCCCAACAATTCCTCTATCTCATATATATATGCTTAAATTAATTACTTGCAAA 4239
QY 241 ACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATATGCTTAAATTAATTACTTGCAAA 300
DB 4240 ACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATATGCTTAAATTAATTACTTGCAAA 4299
QY 301 ACTTGGCTTGACAAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTT 360

Db 4300 ACTTGCTTTGACAAATGGAATGTAGTACAAATGACACCTGTGGAACTTTGGATTTCAGTTT 4350
 QY 361 CGAGAGAACTTACACCTTCCACTCCACACTCTCTCTGGAAAGCGAGTGAAGTGAAGT 420
 Db 4360 CGAGAGAACTTACACCTTCCACTCCACACTCTCTCTGGAAAGCGAGTGAAGTGAAGT 440
 QY 421 CAGGGCTATCTCTGCTAGAGAGATATCTCCAGCTAAAGCCACAAATCAACCTCTGAGCAT 480
 Db 4420 CAGGGCTATCTCTGCTAGAGAGATATCTCCAGCTAAAGCCACAAATCAACCTCTGAGCAT 4470
 QY 481 ATCAATGAGGCTAGCTAGGCC 501
 Db 4480 ATCAATGAGGCTAGCTAGGCC 4500

RESULT 2

AAZ22642/c
 ID AAZ22641 standard; cDNA; 502 BP.

XX
 AC AAZ22641;

DT 08-DEC-1999 (first entry)

XX Novel cDNA sequence 1 derived from the b2HFLS20W cDNA library.

XX interferon; novel sequence; pathogenic infection; antiviral; cell fate;
 KW cell proliferation; immune disorder; leukemia; wound healing; antibody;
 KW assay; ds.

XX Homo sapiens.

XX WO9945117-A2.

XX 10-SEP-1999.

PF 04-MAR-1999; 99WO-US03727.

FR 04-MAR-1998; 98US-0034878.

FR 20-AUG-1998; 98US-0137348.

XX (HYSEQ) HYSEQ INC.

XX Drmanac RT, Crkvenjakov R, Dickson M, Drmanac S, Labat I;

PI Leshkowitz D, Kita D, Ford C, Muiero J;

XX WPI; 1999-551041/46.

XX New interferon polynucleotides useful for treating disorders involving
 PF pathogenic infection, cell fate and differentiation, inflammation etc.

PS Claim 1; Fig 2; 94pp; English.

XX This a novel sequence isolated from the b2HFLS20W cDNA library, prepared
 CC from human fetal liver-spleen tissue, with another novel sequence
 CC (AAZ22642). The polypeptide sequences corresponding to these nucleic
 CC acid sequences have a striking homology to the interferon family of
 CC polypeptides.

CC The interferon polypeptides may be administered therapeutically to
 CC treat/prevent disorders involving pathogenic infection (e.g. viral
 CC and bacterial infections) or cell fate and differentiation (e.g. immune
 CC disorders, leukemias, wound healing, tissue growth etc.). Interferons
 CC are known to have antiviral activity and are believed to be involved in
 CC regulatory processes such as cell proliferation and differentiation.
 CC The polypeptides may also inhibit platelet aggregation and therefore be
 CC used to treat/prevent conditions involving inflammatory responses. They
 CC can be combined with a carrier in compositions which can be administered
 CC to treat disorders as above, and used to generate antibodies useful in
 CC polypeptide detection and purification. They can also be used to
 CC identify binding compounds, by contacting with a compound and detecting
 CC complex formation, optionally within a cell in which the complex drives
 CC expression of a reporter gene. They (and the polynucleotides) are
 CC therefore useful in drug screening assays e.g. for the above disorders.

CC The polynucleotides are also useful to detect polynucleotides encoding
 CC the polypeptides e.g. to diagnose disorders (or susceptibility to
 CC disorders) relating to polypeptide expression, or mark polypeptide-
 CC expressing tissues. They may also be used to produce probes useful to
 CC detect related sequences or for gene mapping.

XX Sequence 502 BP; 141 A; 97 C; 143 G; 118 T; 3 other;

XX Query Match 16.5%; Score 82.8; DB 20; Length 502;

XX Best Local Similarity 63.8%; Pred. No. 2.1e-14;

XX Matches 157; Conservative 0; Mismatches 87; Indels 2; Gaps 2;

QY 203 ATGCTCATATATATGCCCCCTTGGCAATGTGACTTTGCTACTTCTCT-ATCAAGATGTGGA 261
 Db 347 ATTCCTCCCATCTCTGCTCCCACTACATGTGACTTTGCTGCTTCTGTCATTAGAGATGGA 288

QY 282 GCTTATTTTCCCATATATTGCACTAGAGTGGCTTGGCTTGTGCTTTGA-CAATGGAAAT 320
 Db 287 GTCTGTTTCCCACTCTCTTGAATCTGGGCTGATCTTATGATTTGTTTGACCAAGTAAAT 228

QY 321 GTAGTACAAATGACACTGTGGAACTTTGGATTTTAGGTTTGGAGAGAACTTACACCTCC 380
 Db 227 CTGTTGGAAGGATGCTGGCAACTTCTGATGTAGGCTCAAGAAGCTCGCAACTCC 168

QY 351 ACTCACACTCTCTTGGAAACAGATGCAATGTAAAGAACTCAGGGCTATCTGCTAGAGA 440
 Db 167 ACGTCTATTTTGGAGCTCTTCTGCGGCTCTCTGGAAGAAGCTGGACCACTGCTGAGAGA 108

QY 441 CATATG 446

Db 107 CATGTG 102

RESULT 3

AAZ22642/c

ID AAZ22642 standard; cDNA; 854 BP.

XX AAZ22642;

DT 08-DEC-1999 (first entry)

XX Novel nucleotide sequence IFN-HY from the b2HFLS20W cDNA library.

XX interferon; novel sequence; pathogenic infection; antiviral; cell fate;
 KW cell proliferation; immune disorder; leukemia; wound healing; antibody;
 KW assay; ds.

XX Homo sapiens.

XX WO9945117-A2.

XX 10-SEP-1999.

XX 04-MAR-1999; 99WO-US03727.

XX 04-MAR-1998; 98US-0034878.

XX 20-AUG-1998; 98US-0137348.

XX (HYSEQ) HYSEQ INC.

XX Drmanac RT, Crkvenjakov R, Dickson M, Drmanac S, Labat I;

PI Leshkowitz D, Kita D, Ford C, Muiero J;

XX WPI; 1999-551041/46.

XX P-PSDE; AAY42442.

XX New interferon polynucleotides useful for treating disorders involving
 PF pathogenic infection, cell fate and differentiation, inflammation etc.

PS Claim 1; Fig 2; 94pp; English.

XX This a novel sequence isolated from the b2HFLS20W cDNA library, prepared

PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI: 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
XX
PS Claim 1: SEQ ID NO 246; 32pp - Sequence Listing: German.
XX
XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, aneurysm, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 5239 BP; 1541 A; 54 C; 1235 G; 2309 T; 0 other;

Query Match 12.9%; Score 54.6; DB 24; Length 5239;
Best Local Similarity 51.3%; Pred. No. 1.2e-08;
Matches 174; Conservative 0; Mismatches 164; Indels 1; Gaps 17

QY 62 TTTTATATTCAGCAAAATGGTTGGAAATGCTGTCTTAAATGAGATGCTTACGTCGCG 121
DB 3273 TTTAATACCTTTAAAAAACACGCCCAAAATTTAATATTAATATTTCAATCAATTAACCT 3214

QY 122 TCTGACATGAGGTAGAGGAATTTTACACATATATATGCTGCTTAATTTCTTCAAG 130
DB 3213 TCATTACCTTCTTAATACCAACACCTTAAATATTTATACCTTAAAAAACCTTACAA 3154

QY 182 ATGGCCACACAAATTCCTCTCTATCTCTCATATATATGCGGCTTTGGATGTGACTTGGCTA 241
DB 3153 ATATACCATCAATCTCTCTCTATCTCTATATACCTACGCTTTACCGTATACCTTACCA 3104

QY 242 CTCTCTATCAAGATGSGAGCTTATTTTCCCATATATATGACATGAGTTGGCTTCTGA 301
DB 3093 TTCTCTCTATCAAAATTAATATTTCTATATCTTCCATCTTAATCTTAACTTAACTTAA 3024

QY 302 CTTCCTTTGA-CAATGGAAATGATGACATGACACTGCACTTGGACTTTGAGTTT 350
DB 3033 TTACTTTTACCAATAAATATATACCAAAATATATTTATATATATTTTAAACTTAACT 2974

QY 361 CGACAGAACTTACACCTTCCACTCAGACTCTCTTGGAA 399
DB 2973 CGAAAAACCTTACAACTTATATTTTATATTTCTTCAAAA 2935

RESULT 8
ID AAC74337/c standard; cDNA; 1241 BP.
XX
AC AAC74337;
XX
DT 02-FEB-2001 (first entry)
XX
DE Human secreted protein gene 1 SEQ ID NO:11.
XX
KW Human; secreted protein; diagnosis; cyrostatic; immunosuppressive;
KW nootropic; neuroprotective; antiviral; antiallergic; hepatotropic;
KW antidiabetic; antiinflammatory; antitumor; vulnary; anticonvulsant;
KW antibacterial; antifungal; antiparasitic; cardiant; gene therapy;
KW food additive; preservative; chromosome identification; cancer;
KW female reproductive system disorder; immune disorder; wound healing;
KW cardiovascular disorder; neurological disease; infectious disease;
KW infection; ss.
OS Homo sapiens.
XX

PN WC200058340-A2.
XX
PI 05-OCT-2000.
XX
PF 23-MAR-2000; 2000WO-US07724.
XX
PR 26-MAR-1999; 99US-0126510.
PR 07-JAN-2000; 2000US-0174850.
XX
PA (HUMA-) HUYAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM, Komatsoulis G;
XX
DR WPI: 2000-594638/56.
DR P-PSDE; AAB39402.
XX
PT Fifty nucleic acid molecules encoding human secreted proteins, useful
PT in the prevention, treatment and diagnosis of cancer, immune disorders,
PT cardiovascular disorders and neurological diseases
XX
PS Claim 1: Page 329; 391pp; English.
XX
CC The polynucleotide sequences given in AAC74337 to AAC74396 encode the
CC human secreted proteins given in AAB39402 to AAB39451. AAB39452 to
CC AAB39484 represent human secreted polypeptide sequences and proteins
CC homologous to them, which are given in the exemplification of the present
CC invention. Human secreted proteins have activities based on the tissues
CC and cells the genes are expressed in. Example of activities include:
CC cyrostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
CC antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
CC vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic; and
CC cardiant. The polynucleotides and polypeptides are useful for preventing,
CC treating or ameliorating a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. The polypeptides
CC can also be used as a food additive or preservative to increase or
CC decrease storage capabilities. The polynucleotide are useful for
CC chromosome identification. They are also useful as probes for diagnosing
CC a disorder related to the female reproductive system, particularly breast
CC and/or ovary cancer. They are also useful in the gene therapy of breast
CC and ovarian cancer. Secreted protein nucleic acids, proteins,
CC antibodies, agonists and antagonists are useful in the diagnosis,
CC treatment and prevention of: (a) cancer; (b) immune disorders; (c)
CC cardiovascular disorders; (d) wound healing; (e) neurological diseases,
CC and (f) infectious diseases such as viral, bacterial, fungal and
CC parasitic infections. AAC74328 to AAC74336 and AAB39401 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 1241 BP; 363 A; 266 C; 257 G; 355 T; 0 other;

Query Match 12.6%; Score 63; DB 21; Length 1241;
Best Local Similarity 57.4%; Pred. No. 2e-08;
Matches 118; Conservative 0; Mismatches 55; Indels 2; Gaps 2;

QY 146 TCTACACATAATCTGCTTAATTTACTTGGCAAGATGGCCACAA-CAATTCCTCTAT 204
DB 803 TATTCACATCTCTTAATGTGGAAGTTATTAGCAAGATGGTACAAATTATTTCTCCAT 744

QY 205 COTCATATATATGCCCCCTTTGCAATGTGACTTTGCTACTT-CTCTATCAAGATGTGGAGC 263
DB 743 TTCCATACATATACCACTTTTATATATATGCTCTGCCACTTGGCCCATCAAGATGGAAT 684

QY 264 TTATTTTCCCATATATTGCACTACAGATTTGGCCCTCTGACTTGTGACATGCA 318
DB 683 TTATTTTCCCATATACCAATGATCAGGGGTGCTGATGTAACATGCTTTGGCCCAACCA 629

RESULT 9
AAC05322
ID AAC05322 standard; cDNA; 1479 BP.
XX
AC AAC05322;
XX
DT 17-JUL-2001 (first entry)

PF 01-NOV-2000; 2000WO-US03045.
XX
XX
PR 05-NOV-1999; 99JUS-0163581.
PR 30-JUN-2000; 2000US-0215133.
XX
XX (HUMA-): HUMAN GENOME SCI INC.
XX
XX Ruben SM, Komatsoulis GA, Moore PA, Birse CE, Ni J;
XX
XX WPI; 2001-308778/32.
DR P-PSDB; AAEC1506.
DR
XX
XX New nucleic acid molecules encoding 28 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives.
XX
XX
XX Claim 1; Page 477-478; 562pp; English.
XX
XX
CC AAD05130-AAD05379 represent cDNAs corresponding to 28 human secreted
CC protein genes, and AAEC1436-AAEC1513 represent the proteins they encode.
CC AAEC1514-AAEC1544 represent human secreted protein fragments or variants.
CC
CC The genes and their secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 28 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiogenic disorders, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein-encoding cDNA of the invention.
XX
SQ Sequence 1484 BP; 418 A; 337 C; 300 G; 422 T; 7 other;

Query Match 12.3%; Score 61.4; DP 22; Length 1484;
Best Local Similarity 57.2%; Pred. No. 6.5e-08;
Matches 131; Conservative 0; Mismatches 96; Indels 2; Gaps 1;

QY 229 TGTGACTTGTCTACTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTCGACTAGA 288
DB 632 TATGTGTGTGCCACCCCTCCCATCAAGACATGGTGGCTATTTTTCATCCTTTGTCTGG 691
QY 289 GTTGGCCTTCTGACTTGTGACATGGAATGTAGTACAAATGACACTGTGCACTTTG 348
DB 692 GCTGGCCTTGTGTCCATTTGACCAATAGGTATGGTAGAATGGCAATCTGGGACTTCT 751
QY 349 GATTTTAGTTCGAGAGAACTTACACCTTCCACTACACTCTCTTGGAAACAGATGCA 408
DB 752 GAGCCCAACCTT--AAAGACATTTGAACCTTTTGTCTAGTGAGAGGGGAATCAGATACC 809
QY 409 ATGTAAGAGATCAGGGCTATCTCTGCTAGAGACATATGTCCAGCTAAT 457
DB 810 CTGTAAGAGGCTGTGTGTCTCTCTTCTGAAGACAGAGAGGCTCTCTGGAAGAT 858

RESULT 11
AAS46911

ID AAS46911 standard; cDNA; 670 BP.
XX
XX AC AAS46911;
XX
XX DT 19-DEC-2001 (first entry);
XX
XX DE Human G protein-coupled receptor (GPCR) cDNA #93.
XX
XX KW Human; G protein-coupled receptor; GPCR; mental disorder; schizophrenia;
KW neurological disorder; metabolic disorder; cancer; rheumatoid arthritis;
KW thyroid disorder; neurodegenerative disorder; cardiovascular disorder;
KW renal failure; autoimmune disorder; hyperproliferative disorder; HIV; ss;
KW human immunodeficiency virus; viral infection; neuroprotective;
KW immunostimulant; neuroleptic; nootropic; tranquiliser; antidepressant;
KW anorectic; gene therapy.
XX
XX OS Homo sapiens.
XX
XX PN WO200168859-A2.
XX
XX PD 20-SEP-2001.
XX
XX PE 16-MAR-2001; 2001WO-US08456.
XX
XX PR 16-MAR-2000; 2000US-187783P.
PR 16-MAR-2000; 2000US-189907P.
PR 16-MAR-2000; 2000US-189917P.
PR 16-MAR-2000; 2000US-189918P.
PR 16-MAR-2000; 2000US-189960P.
PR 29-MAR-2000; 2000US-192155P.
PR 29-MAR-2000; 2000US-192234P.
PR 29-MAR-2000; 2000US-192830P.
PR 29-MAR-2000; 2000US-192916P.
PR 29-MAR-2000; 2000US-192923P.
PR 29-MAR-2000; 2000US-192933P.
PR 29-MAR-2000; 2000US-192945P.
XX
XX (PNA : PHARMACIA & UPJOHN CO.
XX
XX Vogel G;
XX
XX WPI; 2001-607458/69.
XX P-PSDB; AAU29472.
XX
XX Nucleic acid encoding G protein-coupled receptors, useful for the
XX prevention, diagnosis and treatment of mental disorders.
XX
XX Claim 4; Page 90; 274pp; English.
XX
XX Sequences AAS46819-AAS46946 represent cDNA molecules encoding human G
XX protein-coupled receptor (GPCR) polypeptides. The protein and DNA
XX sequences of the invention can be used to identify compounds which bind
XX to GPCR polypeptides and in screening for compounds that modulate GPCR
XX activity. By screening a human subject for the presence of mutations in
XX GPCR DNA, a GPCR-related disorder or a genetic predisposition can be
XX diagnosed. The sequences can also be used for treatment and prevention of
XX mental disorders such as schizophrenia, neurological disorders such as
XX manic depression, metabolic disorders such as obesity, cancer, rheumatoid
XX arthritis, thyroid disorders such as myxoedema, neurodegenerative
XX disorders such as Parkinson's disease, cardiovascular disorders such as
XX atherosclerosis, renal failure, autoimmune disorders, hyperproliferative
XX disorders such as psoriasis and viral infections such as those caused by
XX HIV.
XX
SQ Sequence 670 BP; 218 A; 134 C; 143 G; 175 T; 0 other;

Query Match 12.2%; Score 61; DP 22; Length 670;
Best Local Similarity 60.6%; Pred. No. 6.4e-08;
Matches 117; Conservative 0; Mismatches 75; Indels 1; Gaps 1;

QY 237 TGTACTTCTCTATCAAGATGTGGAGCTTATTTTCCCATATATTCGACTAGATTCGCT 296
DB 319 TGCAGTGTTCATCAAGAGCTGCAGTCTATTTCCCAACCCCAAAATCCA-AGGTGGCT 377

```

QY 297 TCTGACTTGCTTGGACAAATGGAATGTAGTACAAATGACACTTGGCACTTTGGAACTTTGGATTATTAG 256
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 378 TGTGACTTGCTTTGAAATGAATGAGTGGAGGAATTTTGTGACTTTCCAGACTAG 437
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 357 GTTTCGAGAGAACTTACACCTTCCACTCCACACTCTCTCTTGGAAACCAGATGGAATGTAAAG 416
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 GCTGCAAGAGACTTTGGAGATTCCTCCACTTCCACATTTCTTGGAAATGCTGCCCTGAGATAGCC 497
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 417 AAGTCAGGGCTAT 429
   ||| |||
Db 496 ATGCAAGGAAGAT 510
   ||| |||

RESULT 12
ABK61689
ID ABK61689 standard; DNA; 670 BP.
XX
AC ABK61689;
XX
DT 13-AUG-2002 (first entry);
XX
DE cDNA encoding novel G protein coupled receptor (nGPCR-x) #93.
XX
KW G protein coupled receptor; nGPCR-x; immune response; thyroid disorder;
KW mental disorder; thyrotoxicosis; myxedema; inflammatory condition;
KW Crohn's disease; cell differentiation; homeostasis; rheumatoid arthritis;
KW renal failure; autoimmune disorder; movement disorder; CNS disorder;
KW viral infection; human immunodeficiency virus; HIV; metabolic disorder;
KW cardiovascular disorder; diabetes; obesity; anorexia; cardiomyopathy;
KW proliferative disease; cancer; psoriasis; lung cancer; hormonal disorder;
KW sexual dysfunction; gene; ss.
XX
OS Homo sapiens.
XX
PN US2002056306-A1.
XX
PC 16-MAR-2002.
XX
PF 16-MAR-2001; 2001US-0811284.
XX
PR 16-MAR-2000; 2003US-189781P.
PR 16-MAR-2000; 2003US-189937P.
PR 16-MAR-2000; 2003US-189917P.
PR 16-MAR-2000; 2003US-189913P.
PR 16-MAR-2000; 2003US-189960P.
PR 24-MAR-2000; 2003US-192155P.
PR 27-MAR-2000; 2003US-192234P.
PR 29-MAR-2000; 2003US-192830P.
PR 29-MAR-2000; 2003US-192945P.
PR 29-MAR-2000; 2003US-192945P.
PR 29-MAR-2000; 2003US-192945P.
PR 29-MAR-2000; 2003US-192830P.
PR 29-MAR-2000; 2003US-192945P.
PR 29-MAR-2000; 2003US-192830P.
PR 29-MAR-2000; 2003US-192945P.
XX
PA (VOGE/) VOGEL; G.
XX
XX Vogel; G.
XX
XX WP1; 2002-434856/46.
XX
XX P-PSDB; ABG60760.
XX
XX
XX New isolated nucleic acid encoding a G protein coupled receptor for
XX producing the receptor which can induce an immune response in a mammal
XX
XX
XX Claim 4; Page 65-66; 216pp; English.
XX
XX The invention describes an isolated nucleic acid (I) comprising a

```

```

CC sequence encoding a portion of a G protein coupled receptor (nGPCR-x).
CC (i) is used to produce a recombinant nGPCR-x polypeptide. A polypeptide
CC encoded by (i) is used to induce an immune response in a mammal. nGPCR-x
CC is used to identify a compound that binds to it and/or modulates its
CC activity. (i) is used to identify animal homologues of nGPCR-x. (i) can
CC be used to diagnose a human subject as having a brain or genetic
CC predisposition disorder, such as a mental disorder. (i) is used to
CC screen for an nGPCR-x related disorder including thyroid disorders (e.g.
CC thyrotoxicosis, myxedema), renal failure, inflammatory conditions (e.g.
CC Crohn's disease), diseases related to cell differentiation and
CC homeostasis, rheumatoid arthritis, autoimmune disorders, movement
CC disorders, CNS disorders, viral infections (e.g. Human immunodeficiency
CC virus), metabolic and cardiovascular disorders (e.g. diabetes, obesity,
CC anorexia, cardiomyopathy), proliferative diseases and cancers (e.g.
CC psoriasis, lung cancer), hormonal disorders, sexual dysfunction and
CC hereditary mental disorders in a human patient. A host cell comprising
CC (i) is used to screen for a modulator of nGPCR-x activity. nGPCR-x is
CC used to identify compounds that can treat mental disorders. The
CC polypeptide encoded by (i) is used to purify a G protein from a sample.
CC This sequence encodes a novel G protein coupled receptor (nGPCR-x)
CC protein described in the invention.
XX
SQ Sequence 670 BP; 216 A; 134 C; 143 G; 175 T; 0 other;

```

```

Query Match 12.2%; Score 61; DB 24; Length 670;
Best Local Similarity 60.6%; Pred. No. 6.4e-08;
Matches 117; Conservative C; Mismatches 75; Indels 1; Gaps 1;
QY 237 TGTACTTCTCTATCAAGATGTGGAGCTTATTTCCCATATATTGCATAGAGTTGGCCT 296
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 319 TGCAGTGTTTTCATCAAGAGCTGCAGTCTATTTCCCAACCCCAAAATCCAAGGTGGCCT 377
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 297 TGTGACTTGCTTTGACAAATGGAATGTAGTACAAATGACACTTGTGCACTTTGGATTATTAG 356
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 378 TGTGACTTGCTTTGAAATGAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 437
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 357 GTTTCGAGAGAACTTACACCTTCCACTCCACACTCTCTCTTGGAAACCAGATGGAATGTAAAG 416
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 438 GCTGCAAGAGACTTTGGAGATTCCTCCACTTCCACATTTCTTGGAAATGCTGCCCTGAGATAGCC 497
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 417 AAGTCAGGGCTAT 429
   ||| |||
Db 496 ATGCAAGGAAGAT 510
   ||| |||

```

RESULT 13

ABL32272

ID ABL32272 standard; DNA; 5239 BP.

XX ABL32272;

XX 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 245.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytostatic; nontropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

XX Homo sapiens.

XX WO200200929-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OX nucleic - nucleic search, using sw model

Run on: October 24, 2003, 12:34:49 ; Search time 31.1642 Seconds
(without alignments)
7052.739 Million cell updates/sec

Title: US-09-830-902-1_COPY_4000_4500

Perfect score: 501

Sequence: 1 ccaaaagtctgggattacag.....cgaatgaggttagtaggac 501

Scoring table: IDENTITY_WJC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 22039156 residues

Total number of hits satisfying chosen parameters: 113956

Minimum DB seq length: 3
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents.NA:
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq:
6: /cgn2_6/ptodata/2/ina/backfiles.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53.6	10.7	1586	US-09-173-581-11	Sequence 11, Appl
2	53.6	10.7	1586	US-09-420-915-11	Sequence 11, Appl
3	53.6	10.7	1858	US-09-387-212-1	Sequence 1, Appl
4	53.6	10.7	1869	US-09-948-802-1	Sequence 1, Appl
5	53.6	10.7	2142	US-09-441-039-3	Sequence 1, Appl
6	52.2	10.4	2141	US-09-441-039-1	Sequence 1, Appl
7	47.8	9.5	537	US-08-642-274D-38	Sequence 38, Appl
8	47.8	9.5	537	US-08-952-014C-38	Sequence 39, Appl
9	46.6	9.3	465	US-09-499-522-3	Sequence 5, Appl
10	45.6	9.1	6388	US-09-873-404-3	Sequence 3, Appl
11	45.2	9.0	162450	US-09-345-882-1	Sequence 1, Appl
12	45	9.0	3609	US-09-703-299-11	Sequence 1, Appl
13	44.4	8.9	43950	US-09-735-934A-3	Sequence 3, Appl
14	44.4	8.9	43950	US-10-060-332-3	Sequence 3, Appl
15	44	8.8	17041	US-08-076-011-1	Sequence 1, Appl
16	43.9	8.7	3350	US-09-110-116-2	Sequence 1, Appl
17	43.8	8.7	17138	US-09-813-819-3	Sequence 3, Appl
18	43.8	8.7	17138	US-09-920-048-3	Sequence 3, Appl
19	43.8	8.7	64467	US-09-803-671B-3	Sequence 3, Appl
20	43.6	8.7	2373	US-09-975-762-45	Sequence 45, Appl
21	43.6	8.7	2373	US-09-295-028-45	Sequence 45, Appl
22	43.6	8.7	2373	US-09-106-582-45	Sequence 45, Appl
23	43.2	8.6	46718	US-09-816-093-3	Sequence 3, Appl
24	43	8.6	41694	US-09-536-059-1	Sequence 1, Appl
25	43	8.6	98844	US-09-791-211-10	Sequence 10, Appl
26	42.8	8.5	99500	US-09-798-096-10	Sequence 10, Appl
27	42.6	8.5	14753	US-09-821-736-3	Sequence 3, Appl

C	28	42.2	8.4	35060	3	US-08-814-095-7	Sequence 7, Appl
	29	42.2	8.4	40000	4	US-09-780-049-18	Sequence 18, Appl
	30	42.2	8.4	65042	4	US-09-784-316-3	Sequence 3, Appl
C	31	42.2	8.4	84495	4	US-09-797-906-3	Sequence 3, Appl
	32	42	8.4	776	4	US-09-535-008-37	Sequence 37, Appl
	33	42	8.4	46718	4	US-09-816-093-3	Sequence 3, Appl
C	34	42	8.4	174493	4	US-09-804-471A-3	Sequence 3, Appl
	35	41.8	8.3	36651	4	US-09-738-894A-3	Sequence 3, Appl
	36	41.8	8.3	36651	4	US-09-364-469-3	Sequence 3, Appl
C	37	41.8	8.3	319608	4	US-09-539-333D-1	Sequence 1, Appl
C	38	41.8	8.3	319608	4	US-09-679-409-1	Sequence 1, Appl
	39	41.6	8.3	1811	1	US-08-848-252-1	Sequence 1, Appl
C	40	41.6	8.3	36651	4	US-09-738-894A-3	Sequence 3, Appl
C	41	41.6	8.3	36651	4	US-09-964-469-3	Sequence 3, Appl
	42	41.6	8.3	174493	4	US-09-804-471A-3	Sequence 3, Appl
	43	41.4	8.3	4823	2	US-08-457-254-5	Sequence 5, Appl
	44	41.4	8.3	4823	2	US-08-484-257-20	Sequence 20, Appl
	45	41.4	8.3	4823	3	US-09-999-927-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-09-173-581-11
Sequence 11, Application US/09173581A
Patent No. 6013455
GENERAL INFORMATION:
APPLICANT: Bandmar, Olga
APPLICANT: Tang, Y. Tom
APPLICANT: Hillman, Jennifer L.
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
APPLICANT: Gorgone, Gina
APPLICANT: Azimzai, Valda
APPLICANT: Lu, Aina
TITLE OF INVENTION: Protein Kinase Homologs
FILE REFERENCE: PF-0614 US
CURRENT APPLICATION NUMBER: US/09/173,581A
CURRENT FILING DATE: 1998-10-15
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PERL Program
SEQ ID NO 11
LENGTH: 1586
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: 132750
US-09-173-581-11

Query Match	10.7%	Score 53.6	DB 3	Length 1586
Best Local Similarity	59.2%	Pred. No. 5.7e-07		
Matches 109	Conservative 0	Mismatches 74	Indels 1	Gaps 1
QV	192	CAATTCCTCTATCCTCATATATATGCCCCCTTTGCAATGTGACATTGCTACTTCTATC	251	
DB	1154	CAATGTATCTTCTGGTATCCACACCTCTTGGCAATGAAATTTGCAGCTCTCTCCCTTC	1213	
QV	252	AAGATGTGGAGCTTATTTCCCATATATTGCACATGAGTTGGCCTTCTGACTTGTGTA	311	
DB	1214	CATAAATGAAGTCTCTTCCCCACCATTTGAATCTGGCTGGCACTGTGACTTGTGA	1273	
QV	312	CAATGGAATGTAGTACAAATGACACTGTGCAACTTGGATTGGTTTAGGTTTCGAGAGAACT	370	
DB	1274	TCAATAGAATGTGGAAGAAGTGACTGTATGCCAGTTCGAAGCCTAGGTTTCAAGAGGCCT	1333	
QV	371	TACA 374		
DB	1334	TATA 1337		
RESULT 2				

US-09-420-915-11
; Sequence 11, Application US/09420915
; Patent No. 6264947
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Tang, Y. Tom
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Gorgone, Gira
; APPLICANT: Azimzai, Valda
; APPLICANT: Lu, Aina
; TITLE OF INVENTION: Protein Kinase Homologs
; FILE REFERENCE: PF-0614 US
; CURRENT APPLICATION NUMBER: US/09/420,915
; CURRENT FILING DATE: 1999-10-20
; EARLIER APPLICATION NUMBER: US 09/173,581
; EARLIER FILING DATE: 1998-10-15
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PERL Program
; SEQ ID NO 11
; LENGTH: 1586
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: 132750
US-09-420-915-11

Query Match 10.7%; Score 53.6; DB 3; Length 1586;
Best Local Similarity 59.2%; Pred. No. 6.1e-07;
Matches 109; Conservative 0; Mismatches 74; Indels 1; Gaps 1;

QY 192 CAATTCCTCTATCTCATATATATGCCCCCTTGCAGTGTGACTTGTCTCTATC 251
DB 1154 CAAATGATATCTTTCTGGTATCCACACCTCTTGGCAATGAAATTTGCAGTCTCTCTC 1213

QY 252 AAGATGTGGAGCTTATTTTCCCATATATTTGCAGTGTGACTTGTCTCTATC 311
DB 1214 CATAAATGAAGTCTCTTTCCCCACCCTTTGAAATCTGGGCTGGTGTGACTGTGACTTGA 1273

QY 312 -CAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTGGATTTTCGAGAGAACT 370
DB 1274 TCAATAGAAATGTGGAAGAAGTGTATGCCAGTTCCAAAGCTAGTGTTCAGAGAGGCT 1333

QY 371 TACA 374
DB 1334 TATA 1337

RESULT 3
US-09-387-212-1
; Sequence 1, Application US/09387212A
; Patent No. 6309849
; GENERAL INFORMATION:
; APPLICANT: ROBISON, KEITH E.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN KINASE AND
; TITLE OF INVENTION: PHOSPHATASE HOMOLOGUES AND USES THEREFOR
; FILE REFERENCE: MNI-09C
; CURRENT APPLICATION NUMBER: US/09/387,212A
; CURRENT FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1868
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: All occurrences of n indicate any nucleotide
US-09-387-212-1

Query Match 10.7%; Score 53.6; DB 4; Length 1868;
Best Local Similarity 59.2%; Pred. No. 6.1e-07;

Matches 109; Conservative 0; Mismatches 74; Indels 1; Gaps 1;

QY 192 CAATTCCTCTATCTCATATATATGCCCCCTTGCAGTGTGACTTGTCTCTATC 251
DB 1415 CAAATGATATCTTTCTGGTATCCACACCTCTTGGCAATGAAATTTGCAGTCTCTCTC 1474

QY 252 AAGATGTGGAGCTTATTTTCCCATATATTTGCAGTGTGACTTGTCTCTATC 311
DB 1475 CATAAATGAAGTCTCTTTCCCCACCCTTTGAAATCTGGGCTGGTGTGACTGTGACTTGA 1534

QY 312 -CAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTGGATTTTCGAGAGAACT 370
DB 1535 TCAATAGAAATGTGGAAGAAGTGTATGCCAGTTCCAAAGCTAGTGTTCAGAGAGGCT 1594

QY 371 TACA 374
DB 1595 TATA 1598

RESULT 4
US-09-948-802-1
; Sequence 1, Application US/09948802
; Patent No. 6465232
; GENERAL INFORMATION:
; APPLICANT: ROBISON, KEITH E.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN KINASE AND
; TITLE OF INVENTION: PHOSPHATASE HOMOLOGUES AND USES THEREFOR
; FILE REFERENCE: MNI-09C
; CURRENT APPLICATION NUMBER: US/09/948,802
; CURRENT FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: 09/357,212
; PRIOR FILING DATE: 1999-08-31
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1868
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: All occurrences of n indicate any nucleotide
US-09-948-802-1

Query Match 10.7%; Score 53.6; DB 4; Length 1868;
Best Local Similarity 59.2%; Pred. No. 6.1e-07;
Matches 109; Conservative 0; Mismatches 74; Indels 1; Gaps 1;

QY 192 CAATTCCTCTATCTCATATATATGCCCCCTTGCAGTGTGACTTGTCTCTATC 251
DB 1415 CAAATGATATCTTTCTGGTATCCACACCTCTTGGCAATGAAATTTGCAGTCTCTCTC 1474

QY 252 AAGATGTGGAGCTTATTTTCCCATATATTTGCAGTGTGACTTGTCTCTATC 311
DB 1475 CATAAATGAAGTCTCTTTCCCCACCCTTTGAAATCTGGGCTGGTGTGACTGTGACTTGA 1534

QY 312 -CAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTGGATTTTCGAGAGAACT 370
DB 1535 TCAATAGAAATGTGGAAGAAGTGTATGCCAGTTCCAAAGCTAGTGTTCAGAGAGGCT 1594

QY 371 TACA 374
DB 1595 TATA 1598

RESULT 5
US-09-441-039-3
; Sequence 3, Application US/09441039
; Patent No. 6428980
; GENERAL INFORMATION:
; APPLICANT: Luo, Ying
; APPLICANT: Yu, Pei Wen
; APPLICANT: Shen, Mary
; APPLICANT: Huang, Chao Bai
; TITLE OF INVENTION: No. 6428980el R1P3 Associated Cell Cycle Proteins, Compositions

TITLE OF INVENTION: and Methods of Use
FILE REFERENCE: A68411/DJB/RMS/DAV
CURRENT APPLICATION NUMBER: US/09/441,039
CURRENT FILING DATE: 1999-11-16
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 2142
TYPE: DNA
ORGANISM: Homo sapiens
US-09-441-039-3

Query Match: 10.7%; Score 53.6; DB 4; Length 2142;
Best Local Similarity 59.2%; Pred. No. 6.4e-07;
Matches 109; Conservative 0; Mismatches 74; Indels 1; Gaps 1;
QY 192 CAATTCCCTCCATCCTCATATATATGCCCCCTTGGCAATGTGACTTTGCTACTTCTCATC 251
DB 1698 CAATGTATCTTCTGGTATCCACACCTCTTGGCAATGCAATTTGCAGCTCCTCCCTTC 1757
QY 252 AAGATGTGGAGCTTATTTTCCCATATATTTGCACTAGAGTTGGCTTCTGACTTTGTTGA 311
DB 1758 CATAAATGAAGTCTCTTTCCCAACCAATTTGAATCTGGGTGGCACTGTGACTTGAATTGA 1817
QY 312 -CAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTGGAGAGAACT 370
DB 1818 TCAATAGAAATGTGGAAGAAGTGAAGTGTATGCAAGTTCCAAAGCTAGGTTTCAAGAGGGCT 1877
QY 371 TACA 374
DB 1878 TATA 1881

RESULT 6

US-09-441-039-1
Sequence 1, Application US/09441039
Patent No. 6428980
GENERAL INFORMATION:
APPLICANT: Luc, Ying
APPLICANT: Yu, Pei Wen
APPLICANT: Shen, Mary
APPLICANT: Huang, Chao Bai
TITLE OF INVENTION: No. 6428980el RIP3 Associated Cell Cycle Proteins, Compositions
TITLE OF INVENTION: and Methods of Use
FILE REFERENCE: A68411/DJB/RMS/DAV
CURRENT APPLICATION NUMBER: US/09/441,039
CURRENT FILING DATE: 1999-11-16
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 2141
TYPE: DNA
ORGANISM: Homo sapiens
US-09-441-039-1

Query Match: 10.4%; Score 52.2; DB 4; Length 2141;
Best Local Similarity 62.2%; Pred. No. 1.7e-06;
Matches 115; Conservative 0; Mismatches 68; Indels 2; Gaps 2;
QY 192 CAATTCCCTCCATCCTCATATATATGCCCCCTTGGCAATGTGACTTTGCTACTTCTCAT 250
DB 1696 CAATGTATCTTCTGGTATCCACACCTCTTGGCAATGCAATTTGCAGCTCCTCCCTTC 1755
QY 251 CAAGATGTGGAGCTTATTTTCCCATATATTTGCACTAGAGTTGGCTTCTGACTTTGTTG 310
DB 1756 CATAAATGAAGTCTCTTTCCCAACCAATTTGAATCTGGGTGGCACTGTGACTTGAATTG 1815
QY 311 A-CAATGGAATGTAGTACAAATGACACTGTGCAACTTTGGATTTTAGGTTTGGAGAGAAC 369
DB 1816 ATCAATAGAAATGTGGAAGAAGTGAAGTGTATGCCAGTTCCAAAGCTAGGTTTCAAGAGGCC 1875
QY 370 TTACA 374
DB 1878 TATA 1881

Db 1876 TTATA 1880

RESULT 7

US-08-642-274D-38
Sequence 38, Application US/08642274D
Patent No. 6200749
GENERAL INFORMATION:
APPLICANT: Shiloh, Yosef
TITLE OF INVENTION: MUTATED FORMS OF THE ATAXIA-TELANGIECTASIA GENE AND METHOD TO
TITLE OF INVENTION: SCREEN FOR A PARTIAL A-T PHENOTYPE
FILE REFERENCE: 229000033
CURRENT APPLICATION NUMBER: US/08/642,274D
CURRENT FILING DATE: 1996-05-03
NUMBER OF SEQ ID NOS: 220
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 38
LENGTH: 537
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: genomic
US-08-642-274D-38

Query Match: 9.5%; Score 47.8; DB 3; Length 537;
Best Local Similarity 70.3%; Pred. No. 2.3e-05;
Matches 64; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 1 CCAAGTGTCTGGATTACAGGGTTAGCCACAGCATCCAGGCTTAAGCATGTTAATTAAG 60
DB 12 CCAAGTGTCTGGATTACAGTCTGTAGCCACCGCACTCGGCTTAAGGTTAATTTCTTGA 71

QY 61 TTTTATAATTCAGCAAAATGGTTGGAAAT 91

DB 72 GTACAGAAAACAGCATTATAGTTTGGAAAT 102

RESULT 8

US-08-952-014C-39
Sequence 38, Application US/08952014C
Patent No. 6265158
GENERAL INFORMATION:
APPLICANT: Shiloh, Yosef
TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE AND ITS
TITLE OF INVENTION: GENOMIC ORGANIZATION
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kohn & Associates
STREET: 30500 No. 6265158thwestern Hwy., Suite 410
CITY: Farmington Hills
STATE: Michigan
COUNTRY: U.S.
ZIP: 48334
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DCS/MS-DCS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/952,014C
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Kohn, Kenneth I.
REGISTRATION NUMBER: 30,995
REFERENCE/DOCKET NUMBER: 2290.00028
TELECOMMUNICATION INFORMATION:
TELEPHONE: 810-539-5050
TELEFAX: 810-539-5055
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 537 base pairs
TYPE: nucleic acid

```
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-952-C14C-38

Query Match          9.5%; Score 47.3; DB 3; Length 537;
Best Local Similarity 70.3%; Pred. No. 2.3e-05;
Matches 64; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 1 CCAAGTGTCTGGGATTACAGGCTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAAG 60
DB 12 CCAAGTGTCTGGGATTACAGTGTGTGAGCCACCGCCTTAGGTTAATTCTTGA 71

QY 61 TTTTATTAATTCAGCAAAATGGTTGGAAAT 91
DB 72 GTACAGAAAACAGCATTATAGTTGSAAT 102

RESULT 3
US-09-499-522-5/c
; Sequence 5, Application US/09499522
; Patent No. 6479238
; GENERAL INFORMATION:
; APPLICANT: Blumefeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Bihain, Bernard
; TITLE OF INVENTION: POLYMORPHIC MARKERS OF THE LSR GENE
; FILE REFERENCE: GENSET.053AUS
; CURRENT APPLICATION NUMBER: US/09/499,522
; EARLIER FILING DATE: 2000-02-10
; EARLIER APPLICATION NUMBER: US 60/119,592
; EARLIER FILING DATE: 1999-02-10
; EARLIER APPLICATION NUMBER: US 60/144,784
; EARLIER FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: Patent.pm
; SEQ ID NO 5
; LENGTH: 465
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 126
; OTHER INFORMATION: 99-14417-126 : polymorphic base C or T
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 103..125
; OTHER INFORMATION: 99-14417-126.mis1 potential
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 127..149
; OTHER INFORMATION: 99-14417-126.mis2 potential, complement
; FEATURE:
; NAME/KEY: allele
; LOCATION: 334
; OTHER INFORMATION: 99-14417-334 : polymorphic base C or T
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 311..333
; OTHER INFORMATION: 99-14417-334.mis1 potential
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 335..357
; OTHER INFORMATION: 99-14417-334.mis2 potential
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 1..21
; OTHER INFORMATION: upstream amplification primer
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 447..465
; OTHER INFORMATION: downstream amplification primer, complement
US-09-499-522-5

Query Match          9.3%; Score 45.6; DB 4; Length 465;
Best Local Similarity 59.6%; Pred. No. 5e-05;
Matches 133; Conservative 0; Mismatches 84; Indels 0; Gaps 3;

QY 223 TTGCAATGTGACTTTGCTACT-TCTCTATCAAGATGTGAGAGCTTATTTCCCATATATTG 281
DB 434 TTACAAGGTGACTTTGCCACTCTTCCCATCAAGGATGGAGTCTGTGTCCTCCACCCCTTG 375

QY 282 CACTAGAGTTGGCTTTCTGACTTGTCTT-GACAAATGGAATGTAGTACAAATGACACTGG 340
DB 374 AATGTGGCTGACCTTGTAAATTTCTTTTGGCCAAATCAACATGGTGAAAGTAATGCTGTG 315

QY 341 CAACTTTGGAATTTAGGTTTCGAGAGAACTTACACCTTCCACTCACACTCTCTTTGGAAAC 400
DB 314 TTAGTTTCAAGCTAGACTTCAAGAAACCTTCCGCTTCCATGTACTCTCTTAGTTGACC 255

QY 401 ----CAGATGCAATGTAAAGAGTCAGGGCTATCTCTGCTAGAG 439
DB 254 NTGCCATGACCTTGTGAACATGACCGGGCTAGCTGCTGGAG 212

RESULT 10
US-09-873-404-3
; Sequence 3, Application US/09873404
; Patent No. 6500656
; GENERAL INFORMATION:
; APPLICANT: WEBSTER, Marion et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO01212-CIP
; CURRENT APPLICATION NUMBER: US/09/873,404
; CURRENT FILING DATE: 2001-06-05
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 63589
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)....(63588)
; OTHER INFORMATION: n = A,T,C or G
US-09-873-404-3

Query Match          9.1%; Score 45.6; DB 4; Length 63588;
Best Local Similarity 54.9%; Pred. No. 0.00076;
Matches 90; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1 CCAAGTGTCTGGGATTACAGGCTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAAG 60
DB 417 CCAAGTGTCTGGGATTACAGGCTTAGCCACAGCATCCAGCCTTAAGCATGTTAATTAAAT 4230

QY 61 TTTTATAATTCAGCAAAATGGTTGGAATGTCTTAAATGAGATGCTTAAGCTGCC 120
DB 4231 CTATTGCTGTGTAGCATATTACCCCAAACTTAGTGGCTTAAAGAAATTAACATTATTGTC 4290

QY 121 GTCTGAACATGAGGTAGAAGGAATTTCTACACATAATCATTTGTG 164
DB 4291 TCAGAGATCTGTGTCATCAAGAAATTAGGATGACGATCATTGAG 4334

RESULT 11
US-09-345-882-1/c
; Sequence 1, Application US/09345882
; Patent No. 6399373
; GENERAL INFORMATION:
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: A NUCLEIC ACID ENCODING A RETINOBLASTOMA BINDING PROTEIN (RBP-)
; TITLE OF INVENTION: AND POLYMORPHIC MARKERS ASSOCIATED WITH SAID NUCLEIC ACID.
; FILE REFERENCE: GENSET.031A
; CURRENT APPLICATION NUMBER: US/09/345,882
```

```

; CURRENT FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: US 60/091,315
; PRIOR FILING DATE: 1998-06-30
; PRIOR APPLICATION NUMBER: US 60/111,909
; PRIOR FILING DATE: 1998-12-10
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: Patent.P
; SEQ ID NO :
; LENGTH: 162450
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 72794
; OTHER INFORMATION: 5-124-273 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 93075
; OTHER INFORMATION: 5-127-261 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 90842
; OTHER INFORMATION: 99-1437-325 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 93714
; OTHER INFORMATION: 5-128-60 : polymorphic base deletion of G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 97122
; OTHER INFORMATION: 99-1442-224 : polymorphic base G or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 97152
; OTHER INFORMATION: 5-129-144 : polymorphic base deletion of T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 99098
; OTHER INFORMATION: 5-130-257 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 99117
; OTHER INFORMATION: 5-130-276 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 103806
; OTHER INFORMATION: 5-131-395 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 106940
; OTHER INFORMATION: 5-133-375 : polymorphic base insertion of A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108106
; OTHER INFORMATION: 5-135-155 : polymorphic base insertion of A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108149
; OTHER INFORMATION: 5-135-198 : polymorphic base insertion of G/T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108308
; OTHER INFORMATION: 5-135-357 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 108471
; OTHER INFORMATION: 5-136-174 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 134134
; OTHER INFORMATION: 5-140-120 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 99375...99121
; OTHER INFORMATION: 5-140-348 : polymorphic base insertion of A
; FEATURE:
; NAME/KEY: allele
; LOCATION: 134374
; OTHER INFORMATION: 5-140-361 : polymorphic base insertion of CA
; FEATURE:
; NAME/KEY: allele
; LOCATION: 146328
; OTHER INFORMATION: 5-143-84 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 146345
; OTHER INFORMATION: 5-143-101 : polymorphic base A or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 150329
; OTHER INFORMATION: 5-145-24 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 160031
; OTHER INFORMATION: 5-148-352 : polymorphic base G or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 72771...72817
; OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID30
; FEATURE:
; NAME/KEY: allele
; LOCATION: 72771...72817
; OTHER INFORMATION: polymorphic fragment 5-124-273 SEQ ID51
; FEATURE:
; NAME/KEY: allele
; LOCATION: 88050...88096
; OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID31
; FEATURE:
; NAME/KEY: allele
; LOCATION: 88050...88096
; OTHER INFORMATION: polymorphic fragment 5-127-261 SEQ ID52
; FEATURE:
; NAME/KEY: allele
; LOCATION: 90819...90865
; OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID49
; FEATURE:
; NAME/KEY: allele
; LOCATION: 90819...90865
; OTHER INFORMATION: complement polymorphic fragment 99-1437-325 SEQ ID70
; FEATURE:
; NAME/KEY: allele
; LOCATION: 93690...93736
; OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID32
; FEATURE:
; NAME/KEY: allele
; LOCATION: 93690...93736
; OTHER INFORMATION: polymorphic fragment 5-128-60 SEQ ID53
; FEATURE:
; NAME/KEY: allele
; LOCATION: 97099...97145
; OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID50
; FEATURE:
; NAME/KEY: allele
; LOCATION: 97099...97145
; OTHER INFORMATION: polymorphic fragment 99-1442-224 SEQ ID71
; FEATURE:
; NAME/KEY: allele
; LOCATION: 97130...97177
; OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID33
; FEATURE:
; NAME/KEY: allele
; LOCATION: 97130...97177
; OTHER INFORMATION: polymorphic fragment 5-129-144 SEQ ID54
; FEATURE:
; NAME/KEY: allele
; LOCATION: 99375...99121
```


CURRENT FILING DATE: 2002-02-01
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 3
LENGTH: 43950
TYPE: DNA
ORGANISM: Homo sapiens
US-10-060-332-3

Query Match 8.8% Score 44.4; D5 4; Length 43950;
Best Local Similarity 77.1%; Pred. No. 3.0015;
Matches 54; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 CCAAGTCTGGGATTACAGGGGTTAGCCACAGCATCCAGGCTTAAGCATGTTAATTAG 60
|||
Db 22734 CCAAGTCTGGGATTACAGGGGTTAGCCACAGCATCCAGGCTTAAGCATGTTAATTAG 60
|||
QY 61 TTTTATATAT 20
|||
Db 22734 CTTTACTAT 22803

RESULT 15
US-08-076-011-1
Sequence 1, Application US/0807601;
Patent No. 5521069
GENERAL INFORMATION:
APPLICANT: ONDA, Haruo
APPLICANT: KIMURA, Chiharu
APPLICANT: OKUBO, Shoichi
TITLE OF INVENTION: NOVEL DNA AND USE THEREOF
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN,
ADDRESSEE: ROBERTS & CUSHMAN
STREET: 130 Water Street
CITY: Boston
STATE: Massachusetts
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DCS/MS-DCS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/076,011
FILING DATE: 11-JUN-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/047,246
FILING DATE: 13-APR-1993
NAME:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,676
FILING DATE: 07-AUG-1991
NAME:
ATTORNEY/AGENT INFORMATION:
NAME: RESNICK, DAVID S
REGISTRATION NUMBER: 34235
REFERENCE/DOCKET NUMBER: 41155-CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)523-3400
TELEFAX: (617)523-6440
TELEX: 200291 STRE UR
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704; base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:

NAME/KEY: CDS
LOCATION: join(7540..7850, 9814..9945, 10421..10519,
LOCATION: 11622..11787)
US 08-076-011-1

Query Match 8.8% Score 44; D5 17; Length 17041;
Best Local Similarity 63.0%; Pred. No. 0.0014;
Matches 68; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 1 CCAAGTCTGGGATTACAGGGGTTAGCCACAGCATCCAGGCTTAAGCATGTTAATTAG 60
|||
Db 14792 CCAAGTCTGGGATTACAGGGGTTAGCCACAGCATCCAGGCTTAAGCATGTTAATTAG 60
|||
QY 61 TTTTATATAT 20
|||
Db 14852 TTTTATATAT 20
|||

Search completed: October 24, 2003, 15:42:46
Job time: 34.3542 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:28:54, Search time 133.415 Seconds
(without alignments)
10556.662 Million cell updates/sec

Title: US-09-830-902-1_COPY_4000_4500
Perfect score: 501
Sequence: 1 ccaagtgctggattacg.....tgaatgaggtagtaggac 501

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 1792395 seqs, 1340900451 residues

Total number of hits satisfying chosen parameters: 3584770

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA.*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
- 16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
- 17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68.8	13.7	756	13	US-10-027-632-18943 Sequence 18943, A
2	65.8	13.7	756	13	US-10-027-632-149822 Sequence 149822, A
3	64.6	12.9	586	12	US-10-029-186-12236 Sequence 12236, A
4	64.6	12.9	5239	12	US-10-311-455-245 Sequence 245, App
5	64.6	12.9	367378	12	US-10-312-841-2 Sequence 2, Appli
6	61	12.2	670	9	US-09-811-284-93 Sequence 93, Appl
7	60.6	12.1	624	13	US-10-027-632-55107 Sequence 55107, A
8	60.6	12.1	624	13	US-10-027-632-311174 Sequence 311174, A
9	60.6	12.1	627	13	US-10-027-632-908 Sequence 908, App
10	60.4	12.1	5239	12	US-10-311-455-245 Sequence 245, App
11	60.4	12.1	367378	12	US-10-312-841-1 Sequence 1, Appli
12	60.2	12.0	651	13	US-10-040-739-925 Sequence 925, App
13	59.4	11.9	594	13	US-10-027-632-67207 Sequence 67207, A
14	56.6	11.3	646	13	US-10-027-632-16407 Sequence 16407, A
15	55	11.0	546	13	US-10-027-632-160975 Sequence 160975, A
16	55	11.0	546	13	US-10-027-632-160976 Sequence 160976, A

C 17	55	11.0	546	13	US-10-027-632-160977	Sequence 160977, A
C 18	54.6	10.9	442	13	US-10-027-632-318161	Sequence 318161, A
C 19	54.6	10.9	450	13	US-10-027-632-93145	Sequence 93145, A
C 20	53.6	10.7	590	13	US-10-027-632-85250	Sequence 85250, A
C 21	53.6	10.7	590	13	US-10-027-632-314433	Sequence 314433, A
C 22	53.8	10.7	604	13	US-10-027-632-105017	Sequence 105017, A
C 23	53.8	10.7	624	13	US-10-027-632-214356	Sequence 214356, A
C 24	53.8	10.7	706	13	US-10-027-632-23713	Sequence 13713, A
C 25	53.6	10.7	1586	9	US-09-870-962-11	Sequence 11, Appli
C 26	53.6	10.7	1868	14	US-10-121-925-1	Sequence 1, Appli
C 27	53.6	10.7	2142	24	US-10-165-696-3	Sequence 3, Appli
C 28	53.4	10.7	584	13	US-10-027-632-46655	Sequence 46655, A
C 29	53.4	10.7	584	13	US-10-027-632-295733	Sequence 295733, A
C 30	53.2	10.6	716	13	US-10-027-632-270823	Sequence 270823, A
C 31	53.2	10.6	716	13	US-10-027-632-270824	Sequence 270824, A
C 32	53.2	10.6	752	13	US-10-027-632-174127	Sequence 174127, A
C 33	53.2	10.6	819	13	US-10-027-632-138083	Sequence 138083, A
C 34	52.6	10.5	556	13	US-10-027-632-227143	Sequence 227143, A
C 35	52.6	10.5	397658	10	US-09-813-320-3	Sequence 3, Appli
C 36	52.6	10.4	2141	14	US-10-165-696-1	Sequence 1, Appli
C 37	52	10.4	857	13	US-10-027-632-167271	Sequence 167271, A
C 38	52	10.4	857	13	US-10-027-632-167272	Sequence 167272, A
C 39	52	10.4	857	13	US-10-027-632-167273	Sequence 167273, A
C 40	52	10.4	857	13	US-10-027-632-167274	Sequence 167274, A
C 41	51.6	10.3	1434	10	US-09-764-855-163	Sequence 163, App
C 42	51.6	10.3	1434	14	US-10-072-349-163	Sequence 163, App
C 43	50.6	10.1	702	13	US-10-027-632-137333	Sequence 137333, A
C 44	50.4	10.1	522	13	US-10-027-632-51013	Sequence 51013, A
C 45	50.4	10.1	522	13	US-10-027-632-85803	Sequence 85803, A

ALIGNMENTS

RESULT :
US-10-027-632-18943
Sequence 18943, Application US/10027632
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
PRIOR FILING DATE: 2002-04-10
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 18943
LENGTH: 756
TYPE: DNA
ORGANISM: Human
US-10-027-632-18943

Query Match: 13.7% Score 68.8; DB 13; Length 756;
Best Local Similarity 62.5%; Pred. No. 1e-09;
Matches 140; Conservative 0; Mismatches 82; Indels 2; Gaps 2;
QY 156 ATCATTTGTGCTAAATTTACTTGCAGAGATGGGCAACAAATTCCTCTCATATATA 215
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 109 ATCATTTATGGTGAGTCTTTGAAAAAGTGGGTACAAACAATTCCTCTCTGTGCA 168


```

62  TTTTATAATTCCAGCAAAATGGTTGGAAATGCTGTCTTAAATGAGATGCTTAAGGTGCG 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  TTTANTAACCTTTAAAAAACACCCCAAAATTAAATTAATTTAAATATTCAATCAATTAAC 3214

122  TCTGAACATGAGGTAGAGGAAATCTACACATAATCATGTGCTAAATTAAGTGAAG 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  TCAATACGTTCTAATAACCAACAACCTTAAAAATATTATTACCATTAANAACCTTACAA 3154

182  ATGGCCACACAATTCCTCTATCTCTCATATATATATGCACTTTGCAATGTGACTT 241
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  ATATAACCATCAATCTCTCTATCTCTATATATATAGCTAGCCCTTTACCGTATTAAC 3094

242  CTTCTCTATCAGATGTGGACCTTATTTTCCCATATATATGCACTAGAGTTGGCTT 301
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  TTCTCTCTATCAAAAATAAATTTCTATACCTCTCAATCTTTAAATCTAAACTTATATA 3034

302  CTTGCTTTGA-CAATGGAAATGTASTACAAATGACACCTGTSCAACTTTGAGTTT 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  TTACTTTTAACCAATAAAATATACCAAAATATATATATATATATATATTAACCTTAA 2974

361  CGAGAGAACTTACACCTTCCCTCCTACACCTCTCTTTGGAAA 399
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  CGAAAAACCTTACAACTTATATTTTAAATTTCTCTCAAAAA 2935

RESULT 5
JS-10-312-841-2/c
: Sequence 2, Application US/10312841
: Publication No. US20030186277A1
: GENERAL INFORMATION:
: APPLICANT: Epigenomics AG
: TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
: FILE REFERENCE: E01/1208/WO
: CURRENT APPLICATION NUMBER: US/10/312,841
: CURRENT FILING DATE: 2002-12-30
: NUMBER OF SEQ ID NOS: 2
: SEQ ID NO 2
: LENGTH: 3673778
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
: NAME/KEY: unsure
: LOCATION: (379615)
: US-10-312-841-2

Query Match 12.9%; Score 64.6; DB 12; Length 3673778;
Best Local Similarity 51.3%; Pred. No. 1.5e-06;
Matches 174; Conservative 0; Mismatches 164; Indels 1; Gaps 1;

QY 62 TTTTATAATTCCAGCAAAATGGTTGGAAATGCTGTCTTAAATGAGATGCTTAAGGTGCG 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1463204 TTTANTAACCTTTAAAAAACACCCCAAAATTAAATTAATTTAAATATCTCAATCAATTAANCT 1463145

122  TCTGAACATGAGGTAGAGGAAATCTTACACATAATCATGTGCTAAATTAAGTGAAG 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1463144 TCATTAACGTTCTAATAACCAACAACCTTAAAAATATTTTATTACCATTAANAACCTTACAAA 1463085

182  ATGGCCACACAATTCCTCTATCTCTCATATATATGCACTTTGCAATGTGACTT 241
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1463084 ATATAACCATCAATCTCTCTATCTCTATATATATAGCTAGCCCTTTACCGTATTAACCTTACCA 1463025

242  CTTCTCTATCAGATGTGGACCTTATTTTCCCATATATATGCACTAGAGTTGGCTT 301
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 242 CTTCTCTATCAGATGTGGACCTTATTTTCCCATATATATGCACTAGAGTTGGCTTCTGA 301

1463024 TTCTCTCTATCAAAAATAAATTTCTATACCTTCAATCTTAAATCTTAAACTTATATA 1462965

302  CTTGCTTTGA-CAATGGAAATGTASTACAAATGACACCTGTGCAACTTTGAGTTT 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 302 CTTGCTTTGA-CAATGGAAATGTASTACAAATGACACCTGTGCAACTTTGAGTTT 360

1462964 TTTACTTTAAACCAATAAAATATACCAAAAATAATATATATATAACTTTTAAACTTAAACCT 1462905

361  CGAGAGAACTTACACCTTCCCTCCTACACCTCTCTTTGGAAA 399
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 361 CGAGAGAACTTACACCTTCCCTCCTACACCTCTCTTTGGAAA 399

```

Db 14628304 CGAAAAACCTTACACACTTATATTTAAATTCCCTCAAAAA 1462866

||||| ||||| ||||| ||||| |||||

RESULT 6

US-09-811-284-93

; Sequence 93, Application US/09811284

; Patent No. US20020058306A1

; GENERAL INFORMATION:

; APPLICANT: Vogel, Gabriel

; TITLE OF INVENTION: No. US20020058306A1: G Protein-Coupled Receptors

; FILE REFERENCE: 20167US:

; CURRENT APPLICATION NUMBER: US/09/811,284

; CURRENT FILING DATE: 2001-03-16

; PRIOR APPLICATION NUMBER: 60/189,783

; PRIOR FILING DATE: 2000-03-16

; PRIOR APPLICATION NUMBER: 60/189,907

; PRIOR FILING DATE: 2000-03-16

; PRIOR APPLICATION NUMBER: 60/189,918

; PRIOR FILING DATE: 2000-03-16

; PRIOR APPLICATION NUMBER: 60/189,960

; PRIOR FILING DATE: 2000-03-16

; PRIOR APPLICATION NUMBER: 60/189,917

; PRIOR FILING DATE: 2000-03-16

; PRIOR APPLICATION NUMBER: 60/182,945

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: 60/182,916

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: 60/182,923

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: 60/182,933

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: 60/182,830

; PRIOR FILING DATE: 2000-03-29

; PRIOR APPLICATION NUMBER: 60/182,234

; PRIOR FILING DATE: 2000-03-27

; PRIOR APPLICATION NUMBER: 60/182,155

; PRIOR FILING DATE: 2000-03-24

; PRIOR APPLICATION NUMBER: 60/182,935

; PRIOR FILING DATE: 2000-03-29

; NUMBER OF SEQ ID NOS: 258

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 93

; LENGTH: 670

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-811-284-93

Query Match 12.2%; Score 61; DB 9; Length 670;

Best Local Similarity 60.6%; Pred. No. 1.8e-07;

Matches 117; Conservative 0; Mismatches 75; Index 1; Gap

QY	237	TGCTACTTCTCTATCAAGATGTGGAGCTTATTTCCCATATATTTCCACCTAGACTGTCG
DB	319	TGCAGTGTTTTCATCAAGAGCTGCAGTCTATTTCCACCCCAAAATCCA-AGGTGGC
QY	297	TCTGACTTGTCTTACAAATGAATGTAGTACAAATGACACTGTGCAACTTTTGGATTTT
DB	378	TGTGACTTGTCTTGAATATGAATGCAGTGAAGGAATTTTGTGTGACTTTTCCAGACT
QY	357	GTTTCGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAACCGAGATGCAATGTAA
DB	438	GCTGCAAGAGACTTTGCAGATTTCCCTTCCACATTTCTTGGAAATGCTGCCCTGAGATAG
QY	417	AAGTCAGGGCTAT 429
DB	498	ATGCAAGGAAGAT 510

RESULT 7

US-10-027-632-55107

; Sequence 55107, Application US/10027632

; GENERAL INFORMATION:


```

; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 160975
;   LENGTH: 546
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-160975

Query Match      11.0%; Score 55; DB 13; Length 546;
Best Local Similarity 62.1%; Pred. No. 9,38e-06;
Matches 100; Conservative 3; Mismatches 53; Indels 5; Gaps 1;

QY 206 CTCATATATATGCCCCCTTTGCAATGIGACCTTGCTAGCTGCTATATCAAGATGTGGAGCTT 265
DB | . . . | . . . | . . . | . . . | . . . | . . . | . . . | . . . |
283 CCTGATCAATACCCCTTTGCAATGIGACTTGTAGGTCGTATATCAATATATATATATATAT 229

QY 266 ATTTTCCCATATATGCGACATAGAGTTGGCCCTTCTGACCTTCTGATTTGCAATGGGATGTAGT 325
DB | . . . | . . . | . . . | . . . | . . . | . . . | . . . | . . . |
228 ATTTTCATGCCCTTTGCAATTTGGGCTGGTCTCTGACCTTCTGATTTGCAATGAAATGGCAT 169

QY 326 ACAATGACACTGTGCACTTTGGATTTTACGTTTCGAGAG 366
DB | . . . | . . . | . . . | . . . | . . . | . . . | . . . |
168 AGAAGTGATAGTGAACCAAGTCTCGGTCCTTACAG 128
```

Search completed: October 24, 2003, 18:34:26
Job time : 134.415 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.
OM nucleic - nucleic search, using sw model
Run on: October 24, 2003, 18:12:56 ; Search time 1046.7 Seconds
(without alignment)
11633.509 Million cell updates/sec

Title: US-09-830-902-1_COPY_4000_4500
Perfect score: 501
Sequence: 1ccaaagtgcaggattacag.....tgaatgaggtacgtacgtagccgacg

Scoring table: IDENTITY NJC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152038056 residues
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :				EST:			
1:	em_estba:	1:	em_estba:	1:	em_estba:	1:	em_estba:
2:	em_estba:	2:	em_estba:	2:	em_estba:	2:	em_estba:
3:	em_estba:	3:	em_estba:	3:	em_estba:	3:	em_estba:
4:	em_estba:	4:	em_estba:	4:	em_estba:	4:	em_estba:
5:	em_estba:	5:	em_estba:	5:	em_estba:	5:	em_estba:
6:	em_estba:	6:	em_estba:	6:	em_estba:	6:	em_estba:
7:	em_estba:	7:	em_estba:	7:	em_estba:	7:	em_estba:
8:	em_estba:	8:	em_estba:	8:	em_estba:	8:	em_estba:
9:	gb_est1:	9:	gb_est1:	9:	gb_est1:	9:	gb_est1:
10:	gb_est2:	10:	gb_est2:	10:	gb_est2:	10:	gb_est2:
11:	gb_est3:	11:	gb_est3:	11:	gb_est3:	11:	gb_est3:
12:	gb_est4:	12:	gb_est4:	12:	gb_est4:	12:	gb_est4:
13:	gb_est5:	13:	gb_est5:	13:	gb_est5:	13:	gb_est5:
14:	em_estfun:	14:	em_estfun:	14:	em_estfun:	14:	em_estfun:
15:	em_estfun:	15:	em_estfun:	15:	em_estfun:	15:	em_estfun:
16:	em_estfun:	16:	em_estfun:	16:	em_estfun:	16:	em_estfun:
17:	em_gss_inv:	17:	em_gss_inv:	17:	em_gss_inv:	17:	em_gss_inv:
18:	em_gss_inv:	18:	em_gss_inv:	18:	em_gss_inv:	18:	em_gss_inv:
19:	em_gss_pn:	19:	em_gss_pn:	19:	em_gss_pn:	19:	em_gss_pn:
20:	em_gss_vrt:	20:	em_gss_vrt:	20:	em_gss_vrt:	20:	em_gss_vrt:
21:	em_gss_fun:	21:	em_gss_fun:	21:	em_gss_fun:	21:	em_gss_fun:
22:	em_gss_mam:	22:	em_gss_mam:	22:	em_gss_mam:	22:	em_gss_mam:
23:	em_gss_mus:	23:	em_gss_mus:	23:	em_gss_mus:	23:	em_gss_mus:
24:	em_gss_pro:	24:	em_gss_pro:	24:	em_gss_pro:	24:	em_gss_pro:
25:	em_gss_rod:	25:	em_gss_rod:	25:	em_gss_rod:	25:	em_gss_rod:
26:	em_gss_phg:	26:	em_gss_phg:	26:	em_gss_phg:	26:	em_gss_phg:
27:	em_gss_vrt:	27:	em_gss_vrt:	27:	em_gss_vrt:	27:	em_gss_vrt:
28:	gb_gss1:	28:	gb_gss1:	28:	gb_gss1:	28:	gb_gss1:
29:	gb_gss2:	29:	gb_gss2:	29:	gb_gss2:	29:	gb_gss2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
C 1	95	17.0	726 29 AG030296 Pan troglodytes
C 2	77	15.4	670 29 BZ610078
C 3	76.6	15.3	657 24 CB215585
C 4	73.8	14.7	442 28 BAC0946

5	71	14.2	655	29	AG046887	AG046887 Pan troglodytes
6	70.6	14.1	608	28	AQ351210	AQ351210 RPO111-12
7	67.8	13.5	555	28	AQ614358	AQ614358 HS 5126-B
8	67.4	13.5	722	13	BX103709	BX103709 BX103709
9	66.4	13.3	588	9	AL703889	AL703889 DKP26865
10	65.6	13.2	477	9	A1041146	A1041146 ov69f02.x
11	65	13.0	474	10	BF929327	BF929327 112-NTC20
12	64.9	12.9	401	28	AQ080797	AQ080797 CIT-HSP-2
13	64.6	12.9	1335	11	AK046951	AK046951 Mus musculus
14	63.9	12.7	865	29	CC142611	CC142611 NEL-37D22
15	63.6	12.7	497	13	BX116065	BX116065 BX116065
16	63.4	12.7	518	13	BU656781	BU656781 C116209.2
17	63	12.6	726	28	AQ389497	AQ389497 RPO111-14
18	62.2	12.4	492	28	B29328	B29328 HS-1049-B1-
19	62	12.4	468	28	AQ239410	AQ239410 CIT-HSP-2
20	62	12.4	540	28	AQ175020	AQ175020 HS-321C-A
21	61.4	12.3	526	28	AQ175021	AQ175021 HS-321C-A
22	59.6	11.9	457	28	AQ769797	AQ769797 HS-317C-B
23	59.2	11.8	373	28	B83258	B83258 RPO111-25K1
24	59.2	11.8	498	28	AQ014114	AQ014114 RPO111-25
25	59	11.8	502	2	HSX073396	HSX073396 Homo sapiens
26	59	11.8	529	10	BF435562	BF435562 rac33a02
27	58.6	11.7	510	14	CB270506	CB270506 1009413 H
28	57.6	11.5	1201	9	AJ579387	AJ579387 AJ579387
29	55.8	11.1	680	13	BU685340	BU685340 JI-CF-EU1
30	55	11.0	511	28	AQ894910	AQ894910 HS 3193-A
31	54.2	10.8	421	28	AQ194975	AQ194975 RPO111-57
32	53.6	10.7	453	9	A1334217	A1334217 qg09f12.x
33	53.6	10.7	453	9	AA582672	AA582672 nm53h12.s
34	53.6	10.7	459	14	CB069503	CB069503 1s15h09.y
35	53.6	10.7	461	28	AQ003949	AQ003949 CIT-HSP-2
36	53.6	10.7	469	9	A1573185	A1573185 tr70f04.x
37	53.6	10.7	506	9	AA706818	AA706818 zj30e03.s
38	53.6	10.7	514	13	BU076998	BU076998 1n20a05.x
39	53.6	10.7	556	14	CD369377	CD369377 JI-H-FT1-
40	53.6	10.7	622	14	CA427318	CA427318 JI-H-FT1-
41	53.6	10.7	633	14	CB528830	CB528830 JI-H-FT2-
42	53.6	10.7	666	14	CD370383	CD370383 JI-H-FT1-
43	53.6	10.7	670	13	BQ575071	BQ575071 JI-H-EZ1-
44	53.6	10.7	697	13	BX106138	BX106138 BX106138
45	53.6	10.7	956	13	EX352121	EX352121 EX352121

ALIGNMENTS

RESULT 1				726 bp DNA linear GSS 01-NOV-2001			
AG030296/c				Pan troglodytes DNA, clone: PTB-002K06.R, genomic survey sequence.			
DEFINITION				AG030296			
ACCESSION				AG030296.1			
VERSION				GSS			
KEYWORDS				Pan troglodytes (chimpanzee)			
SOURCE				Pan troglodytes			
ORGANISM				Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.			
REFERENCE				1			
AUTHORS				Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.			
TITLE				BAC end sequences of library P7B			
JOURNAL				Unpublished			
REFERENCE				2			
AUTHORS				Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.			
TITLE				Direct Submission			
JOURNAL				Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan			
COMMENT				E-mail: chimpanzee@gsr.riken.go.jp, URL: http://hgp.gsriken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170			
				Clones are derived from the chimpanzee BAC library P7B. This BAC end was generated during the R&D process and may have higher chance of			

clone tracking errors.

PRIMERS

Sequencing: M13rev

LIBRARY

Vector : pKS145

R.Site 1 : SacI

R.Site 2 : SacI

Location/Qualifiers

i. 726

/organism="Pan troglodytes"

/mol_type="genomic DNA"

/db_xref="taxon:9598"

/clone="P7B-002K06.R"

/sex="male"

/cell_type="lymphoblast"

/clone_lib="P7B Chimpanzee Male RAC library"

196 a 167 c 155 g 207 t 1 others

FEATURES

source

BASE COUNT
ORIGIN

Query Match 17.0%; Score 65; DB 29; Length 726;
Best Local Similarity 62.5%; Pred. No. 0.00034;
Matches 168; Conservative 0; Mismatches 95; Indels 6; Gaps 2;

QY 144 ATTCTACACATAATCATTTGGCTAAATTACTTGCAAAGATGGCCACACACAATTCCTCCTA 203
|||
DB 437 AATTGAAGATGTCCTATACGGTCGATTGTTCCAAGAGCACTGCACACACTCCATCTA 378
|||
QY 204 TCCTCATATATATGCCCCCTTGCACATGTGACCTTGTCTACTTCTCTATACAGATGTGGAGC 263
|||
DB 377 TC-----TTACATGCCCTTTGCAATATGACTTTTCCACTTTTCCATCAGACATAGACT 323
|||
QY 264 TTATTTTCCCATATATTGCACTAGAGTTGGCCCTTCTGACTTGTCTTGA-CAATGGATGT 322
|||
DB 322 CTATCTTACCTCCCTTAAATCTGGACTTATCTTATGAGTTGCTTTAACCAATATCATGT 263
|||
QY 323 AGTACAAATGACACTGTGGCACTTTGGATTGTAGCTTCAGAGAGAAATTACACCTTCAC 392
|||
DB 262 GGTAGAAATGACAATGTGCACACTCCCAAGACAAACTTCAGAAATCCCAACACCTTCAT 203
|||
QY 393 TCACACTCTCTTGGAAACCAAGATGCAATG 411
|||
DB 202 TTTTATGTCTTGGAAATGCACCTGCCATG 174
|||

RESULT 2

BZ610078/c

LOCUS

BZ610078 Human MCF7 breast cancer cell line library (MCF7_1) Homo
sapiens genomic clone MCF7_1-24022, genomic survey sequence.

DEFINITION

BZ610078

VERSION

BZ610078.1 G131518639

KEYWORDS

GSS.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1

AUTHORS

Gray, J.W. and Collins, C.

TITLE

ESP: a sequence-based approach to the structural genomics of tumors

JOURNAL

Unpublished (2002)

COMMENT

Contact: Volix SV

Colin Collins' lab

UCSF Comprehensive Cancer Center

UCSF Box 0808, San Francisco, CA 94143-0808, USA

Tel: 415 502 7066

Fax: 415 502 5665

Email: svolik@cc.ucsf.edu

This clone is available from Amplicon Express

http://www.genomex.com

Class: BAC ends.

Location/Qualifiers

source

i. 670

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/clone="MCF7_1-24022"

/sex="female"

/clone_lib="Human MCF7 breast cancer cell line library

(MCF7_1)"

/note="Vector: pECBAC1; Site 1: HindIII; This library was

constructed from MCF7 breast cancer cell line by Amplicon

Express (http://www.genomex.com) using their standard

procedure."

215 a 159 c 118 g 178 t

ORIGIN

Query Match 15.4%; Score 77; DB 29; Length 670;
Best Local Similarity 61.4%; Pred. No. 0.0063;
Matches 143; Conservative 0; Mismatches 85; Indels 5; Gaps 1;

QY 154 TAATCATTTGTGCTAAATTACTTGCAAAGATGGCCACACACAATTCCTCCTATCCTATATA 213
|||
DB 265 TAGGAAGTGTGTAGATTGCTTGCAAAGATGACCAAAACAATTCCTTTATCCTGCG--- 210
|||
QY 214 TATGCCCTTTGCAATGTGACTTTTGCTACTTCTCTATCAAGATGTGGAGCTTATTTTCC 273
|||
DB 209 -ATGCCCTTTTGCAAGTGTGATTTTGCCAAATTACTATTAGAGGTGGAGTCTACTTCCA 151
|||
QY 274 ATATATTGCACTAGAGTTGGCCTTCTGACTTGTCTTTGACAATGGAATGATGACAAATGA 333
|||
DB 150 ACCCCTTGAAATCTGGGAGGGGCTTGTAATTGCTTTGTCCAATAGAAATGATCGAAGGTG 91
|||
QY 334 CACTGTGCACACTTTGGATTTTAGGTTTCGAGAGAGAACTTACACCTTCCACTCAC 386
|||
DB 90 ACCTCTGTGACTTCTGAGACCAAGAGGTGAGGAGGCGCTTGCAGCTTTCCCGCTC 38
|||

RESULT 3

CB215585

LOCUS

CB215585

DEFINITION

5', mRNA sequence.

ACCESSION

CB215585

VERSION

CB215585.1 G128263777

KEYWORDS

EST.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens

REFERENCE

1

AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

TITLE

Tumor Gene Index

JOURNAL

Unpublished

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs@mail.nih.gov

cDNA Library Preparation:

cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC)

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

info@image.llnl.gov

Plate: LLAM13163 row: L column: 21

Seq primer: M13RP1 reverse primer (ABI).

FEATURES

Location/Qualifiers

i. 657

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5937188"

/sex="female"

/tissue_type="normal endometrium, late proliferative phase

, cycle day 13"

/lab_host="DH10B (TI-resistant)"
/clone_lib="NICH0 HS_Utl1"
/note="Organ: uterus; Vector: PCMV-SFRT6.1.cddb (ResGen, Invitrogen Corporation); Site_1: Not; Site_2: EcorV; Cloned unidirectionally from microquantity amounts of mRNA from normal endometrial tissue (late proliferative phase, cycle day 13). Average insert size 1.9 kb. Library constructed by ResGen (Invitrogen Corporation)."
BASE COUNT 196 a 156 c 125 g 183 t
ORIGIN

Query Match 15.3%; Score 76.6; DP 14; Length 657;
Best Local Similarity 62.3%; Pred. No. 0.0273;
Matches 154; Conservative 0; Mismatches 89; Indels 4; Gaps 2;
QY 196 TCCTCCATTCCTCATATATATGCGCCCTTGGCAATGTGACCTTGTACTTCTCTATCATCAGCA 255
DB 5 TCACCCCTTCCTGGATACACATACATATATATATGCGCTTGTGCTTGTCTTATCATCAAG 64
QY 256 TGTGGAGCTTATTTCCCATATATATGCACTAGAGTGGCTTGGCTTGGCTTGGCTTGGCTTGG 314
DB 65 AGTGGAAATCGACTTACCTAATCTGTAATGTGAGCTGGCTTGGCTTGGCTTGGCTTGGCTTGG 124
QY 315 TGAATGTAGTACAAATGACACTGTGCAACTTTGAGATTTTGGCTTGGAGAGACTTACCA 374
DB 125 TGAATGTGCCAAGCGACCTTGTGCCAATTCGAGTCTAAGCTTCAAGAGGCTTGTG 184
QY 375 CATTCCCA--CTCACACTCTCTTGGAAACCGAGATGCAATGTAAAGAGTCAAGCTATCC 431
DB 185 TGTTCAACTCTCCCTCTTAGAGCGCTGCCAGATACCATGTGTGATGAAGCCCGAGCTAGCC 244
QY 432 TGCTAGA 438
DB 245 TGCTAGA 251

RESULT 4
LOCUS B80948/c
DEFINITION CIT-HSP-2025N4.TPB CIT-HSP Homo sapiens genomic clone 2025N4, genomic survey sequence.
ACCESSION B80948
VERSION B80948.1 GI:2867971
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 442)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., White,C., Shizuya,H., Simon,M. and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)
JOURNAL Unpublished
COMMENT Other GSSs: CIT-HSP-2025N4.TR CIT-HSP-2025N4.TP
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html
Seq primer: M13-21
Class: BAC ends.
Location/Qualifiers
1..442
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7047102"

/db_xref="taxon:9606"
/clone="2025N4"
/sex="Male"
/cell_type="Sperm"
/clone_lib="CIT-HSP"
/note="Vector: pBelcBAC11; Site_1: HindIII; Site_2: HindIII"
BASE COUNT 133 a 111 c 93 g 105 t
ORIGIN

Query Match 14.7%; Score 73.8; DB 28; Length 442;
Best Local Similarity 60.5%; Pred. No. 0.02;
Matches 141; Conservative 0; Mismatches 97; Indels 5; Gaps 1;
QY 154 TAATCATTTGTGCTAATTAATTACTTGGAAAGATGGCCACACAAATTCCTCTATCTCATATA 213
DB 313 TAGAAGTGTGGTAGATGTGCTTGGAAAGATGACCAATCAATTCCTTTATCTCTG----- 264
QY 214 TATGCGCCCTTTGGCAATGTGACTTTGCTACTTCTCTATCAAGATGTGGAGCTTATTTTCCC 273
DB 263 -ATGCTTTTGGCAGTGTGATTTTGGCAATTTACTATTAGAGGTGGAGTCTACTTCCCA 295
QY 274 ATATATTGCACTAGAGTGGCTTCTGACTTSCCTTGACAATGGATGTAGTACAAATGA 333
DB 294 ACCCTTGAAATCTGGAGGGGCTTGTAAATTTGCTTTGCCAATAGATGTATGGAAGGTG 145
QY 334 CACTGTGCAACTTTGGATTTTAGGTTTCGAGAGAGAACTTACACTTCGACTCAG 386
DB 144 ACCTCTGTGACTTCTGAGACCAGAGTGGAGGAGGCGCATGCAGCTTTCCTCCCTG 92

RESULT 5
LOCUS AG046887
DEFINITION Pan troglodytes DNA, clone: PTB-026D22.F, genomic survey sequence.
ACCESSION AG046887
VERSION AG046887.1 GI:16583773
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 655)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel:81-45-503-9111, Fax:81-45-503-9170
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS
Sequencing: -21M13
LIBRARY Vector : pKS145
R.Site 1 : SacI
R.Site 2 : SacI
Location/Qualifiers
1..655
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="PTB-026D22.F"
/sex="male"
/cell_type="lymphoblast"

```

/clone_lib="PTE Chimpanzee Male BAC Library"
BASE COUNT      164 a 147 c 187 g 157 t
CRIGIN

Query Match      14.2%; Score 71; DB 29; Length 655;
Best Local Similarity 70.3%; Pred. No. 0.056;
Matches 109; Conservative 0; Mismatches 45; Indels 1; Gaps 1;

QY 287 GAGTTGGCTTCTGACTTCTTTGA CAATGGAATGAGTACAAATGACACTGTGCAACT 145
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 462 GAGCTGGCCCTGTGACTTGTCTTTGACCAATTAATGAGTAGAAGTGTCTTGTGCTT 521
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 346 TTGGATTTTAGCTTTCGAGAGAACTTACACCTTCCACTCACACTCTCTTGGAAATCAGAT 403
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 522 TCCAGATTAGCTCTTTAGAGAGCCCTTGCAGATTCCATTTTGGCTTCTTGGAGCCAGCT 581
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 406 GCAATGTAAGAAGTCAGGCTATCTCTGCTAGAGA 440
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 582 ACTATGTAATAAGTCTGGACTATCTCTGCTGGA 616
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
AQ351210      608 bp DNA linear GSS 07-MAY-1999
LOCUS
DEFINITION
  RPC11-120H19.TV RPC1-11 Homo sapiens genomic clone RPC1-11-120H19,
  genomic survey sequence.
ACCESSION
  AQ351210
VERSION
  AQ351210.1 GI:4178545
KEYWORDS
  GSS.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 608)
AUTHORS
  Zhao,S., Adams,M.D., Nierman,W., Vales,J., de Jong,P. and Venter,
  J.C.
TITLE
  Use of BAC End Sequences from Library RPC1-11 for Sequence-Ready
  Map Building
JOURNAL
  Unpublished
COMMENT
  Contact: Shaying Zhao, William Nierman, Mark Adams
  Department of Eukaryotic Genomics
  The Institute for Genomic Research
  9712 Medical Center Dr., Rockville, MD 20850
  Tel: 301 838 0200
  Fax: 301 838 0208
  Email: hbest@igrr.org
  Clones are derived from the human BAC library RPC1-11. For BAC
  library availability, please contact Peter de Jong
  (peter@dejong.med.buffalo.edu). Clones may be purchased from
  BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
  Research Genetics (info@resgen.com). BAC end search page:
  http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html
  Seq primer: T7
  Class: BAC ends.
  Location/Qualifiers
    1..608
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="GDB:754582"
      /db_xref="taxon:9606"
      /clone="RPC1-11-120H19"
      /sex="Male"
      /cell_type="Lymphocytes"
      /clone_lib="RPC1-11"
      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
      RPC111 Human Male BAC Library"
BASE COUNT      176 a 146 c 126 g 160 t
ORIGIN

Query Match      14.1%; Score 70.6; DB 28; Length 638;
Best Local Similarity 57.8%; Pred. No. 0.065;
Matches 189; Conservative 0; Mismatches 124; Indels 14; Gaps 3;

```

```

QY 161 TGTGCTAAATTACTTGGCAAGATGGCCACAAATTCCTCTCTATATATATATGCC- 219
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 133 TGTGGTGGATTATTGGCAAGACGGCATCAACAATTCCTCCATATAGGTATGCAT 192
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 220 CTTTGGCAATGTGACTTTCCTACTTCTCTATCAA-----GATGTGGAGCTTATT 268
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 193 COTATGGATGAGACCTTGTAGCTCCTCTTATCAAGCTCCTCTCTTGTAGCTTCTCTTATT 252
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 269 TTCCCATATATTGCACATAGAGTTGGCTTCTGACTTTTGGATTGTTGACAAATGGAAATGATACAA 328
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 253 TTTCACACCTTTAAATTTGGCTGGCTGTGAGTTACTTGGACCCATAGAAATGCRACAGA 312
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 329 AATGACACTGTGCAACTTTGGATTTTAGCTTTTCGAGAGAACTTACACCTTCCACTCACAC 388
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 313 AATGGCATTCGATGACATCCGAGCCGTGGCATCAAGAGGACTTACAGATTCTTCTCTCAC 372
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 389 TCTCTTGGAAA--CCAGATGCAATGTAAAGAAGTCAGGCTATCTCTGAGAGACATATG 446
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 373 CCTCTTAGACAGTGTATGCTACCATGTGGGGAAGCTCAAGCTAGCTGCTGGAGGACAACT 432
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 447 TCCCAGCTAATAGCCACATCAACCTC 473
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 433 TCACATSCAACAGAGATGAGCCAGCCC 459
  ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
AQ614358      555 bp DNA linear GSS 15-JUN-1999
LOCUS
DEFINITION
  HS_5126_B1_G10_77A RPC1-11 Human Male BAC Library Homo sapiens
  genomic clone Plate-702 Col=19 Row=N, genomic survey sequence.
ACCESSION
  AQ614358
VERSION
  AQ614358.1 GI:5075634
KEYWORDS
  GSS.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 555)
AUTHORS
  Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
  Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
  Hood,L.
TITLE
  Sequence-tagged connectors: A sequence approach to mapping and
  scanning the human genome
JOURNAL
  Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE
  99340589
PUBMED
  10449764
COMMENT
  Contact: Mahairas GG, Wallace JC, Hood L
  High Throughput Sequencing Center
  University of Washington
  401 Queen Anne Avenue North, Seattle, WA 98109, USA
  Tel: (206) 616-3618
  Fax: (206) 616-3857
  Email: jwallace@u.washington.edu
  Clones are derived from the human BAC library RPC1-11. For BAC
  library availability, please contact Peter de Jong
  (peter@dejong.med.buffalo.edu). Clones may be purchased from
  BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
  or from Research Genetics (info@resgen.com). BAC end Web Server:
  http://www.husc.washington.edu
  Plate: 702 row: N column: 19
  Seq primer: T7
  Class: BAC ends
  High quality sequence stop: 555.
  Location/Qualifiers
    1..555
      /organism="Homo sapiens"
      /mol_type="genomic DNA"
      /db_xref="taxon:9606"
      /clone="Plate=702 Col=19 Row=N"
      /sex="male"
      /clone_lib="RPC1-11 Human Male BAC Library"
      /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
      Male blood DNA was isolated from one randomly chosen donor

```



```

/db_xref="taxon:9606"
/db_xref="taxon:9606"
/clone_lib="N02022"
/clone_lib="N02022"
/notes="Organ: nervous tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from CRESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research;
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      117 a      102 c      107 g      147 t      1 others
ORIGIN

```

```

Query Match      13.0%; Score 65; DB 10; Length 474;
Best Local Similarity 57.3%; Pred. No. 0.5;
Matches 136; Conservative 0; Mismatches 100; Indels 3; Gaps 1;

QY 234 CTTTGCTACTTCTATCAAGATGTTGGAGCTTATTTTCCATATATATGGCACTAGAGTTGG 293
DB 324 CTCTGCAGCTCTCTATCAAGAGGTGCAATCTCTTTCTCCCAACATGATGCTGGCTGAG 265
QY 294 CTTTCTGACTTGTGACATGGAATGTAGTACAAATGACACTGTGGCACTTTGGATT 353
DB 264 CTTGTGCTTGTATTTGACAGAGACAGAGAGAGAGTATGATATGCAATTCGGGAGAG 205
QY 354 TAGGTTTCGAGAGAACTTACAG---CTTCCACTCAGACAGTCTTTGGAAACAGATGCAAT 410
DB 204 TAAGCTTCAAAAGAGTCTCTCTAGGCTTCCCTCCGACACGGATCTCTCCAGCTCTCT 145
QY 411 GTAAAGAGTTCAGGGCTATCTCTAGAGACATATGTCCAGAGCTATAGGCAACATCAAC 470
DB 144 GTGAAGAGTTCAGGCTGGCTTGTCTGGAGACATGTGGCCAGCTTACAGCAGCACCAC 85
QY 471 C 47;
DB 84 C 84

```

```

RESULT 12
AC080797
LOCUS      AC080797              401 bp      DNA      linear      GSS 20-AUG-1998
DEFINITION C17-HSP-2367H8.TF C17-HSP Homo sapiens genomic clone 2367H8,
genomic survey sequence.
ACCESSION  AC080797
VERSION    AC080797.1  GI:3441981
KEYWORDS   GSS.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 401)
AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linner,K., Golden,K.,
Berry,K., Granger,D., Sub,E., Wible,C., Shizuya,H., Simon,M. and
Venter,J.C.
TITLE     Use of a random human BAC End Sequence Database for Sequence-Ready
Map Building
JOURNAL    Unpublished
COMMENT    Other GSSs: C17-HSP-2367H8.TR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamads@tigr.org
Clones are available from Research Genetics (info@resgen.com). RAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13-21
Class: BAC ends.
FEATURES             Location/Qualifiers
     1..401
     /organism="Homo sapiens"

```

```

/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone_lib="2367H8"
/sex="Male"
/cell_type="Sperm"
/clone_lib="C17-HSP"
/notes="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
HindIII"
BASE COUNT      101 a      83 c      98 g      119 t
ORIGIN

```

```

Query Match      12.9%; Score 64.8; DB 28; Length 401;
Best Local Similarity 63.0%; Pred. No. 0.53;
Matches 116; Conservative 0; Mismatches 67; Indels 1; Gaps 1;

QY 170 TTAATTGCAAGATGGCCACACAAATTCCTCTATCTCTCATATATATGCGGCTTTGCAAT 229
DB 75 TCAATTGCAAGATGGCTCTCAATAATTTCTCCCATGTCTATGGGCAGGTCTTTTTCAGT 134
QY 210 GTGACTTTGGCACT-TCTCTATCAAGATGTGGAGCTTATTTTCCATATATTCACACTAGA 288
DB 135 GTGACTTTGGCACTCTTTCTGTCAAGAAGTAGAGTTTATTTTCCATCTCTTTCACCTAG 194
QY 289 GTTGGCTTGTGACTTGTCTTGACAAATGGATGTAGTACAAATGACACTGTGCAACTTTG 348
DB 135 TCTGGTCTTGTACTTGTCTTATCAAGTAAATGTGTAGAGTACTAGGGGATTCCTGA 254
QY 349 GATT 352
DB 255 GCTT 258

```

```

RESULT 13
AC046951
LOCUS      AC046951              1335 bp      mRNA      linear      HTC 05-DEC-2002
DEFINITION Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length
enriched library, clone:189300C7-11; product:unknown EST, full insert
sequence.
ACCESSION  AC046951
VERSION    AC046951.1  GI:26091826
KEYWORDS   HTC; CAP trapper.
SOURCE     Mus musculus (house mouse)
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1
AUTHORS   Carninci,P. and Hayashizaki,Y.
TITLE     High-efficiency full-length cDNA cloning
JOURNAL    Meth. Enzymol. 303, 19-44 (1999)
MEDLINE    99279253
PubMed     10349636
REFERENCE  2
AUTHORS   Carrinci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL    Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE    20493374
PubMed     11042159
REFERENCE  3
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P.,
Konno,H., Akiyama,G., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
Suzuki,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Futawake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multichannel sequencer
JOURNAL    Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE    20530913
PubMed     11075861
REFERENCE  4

```

AUTHORS Kawai,J., Shitagawa,A., Shibata,K., Yoshino,X., Itoh,X., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,Y., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,T., Saito,T., Okazaki,Y., Gotohori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kueh,P., Lewis,S., Matsuo,Y., Nakado,I., Pesole,G., Quackenbush,C., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,Y., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,X.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,X., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kariya,M., Lee,K.H., Lyons,F., Marchionni,J., Mashima,J., Mazzarelli,J., Montaehts,P., Nordens,P., Ring,B., Ringwald,X., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.P., Suzuki,H., Toyooka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L., Wyrshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohsaki,S. and Hayashizaki,Y.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409 (6821), 685-690 (2001)

MEDLINE 2108566C

PUBMED 11217851

REFERENCE 5

AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation of 50,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

REFERENCE 6 (bases 1 to 1335)

AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,T., Kasukawa,T., Katch,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koyama,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,Y., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shiragawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE Direct Submission

JOURNAL Submitted (16-JUL-2001) Yoshinide Hayashizaki, The Institute of Physiological and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gscl.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL:http://genome.gsc.riken.go.jp/ URL:http://fantom.gsc.riken.go.jp/

FEATURES Location/Qualifiers
 source
 1..1335
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM:DB:B93000711"
 /db_xref="taxon:10090"
 /clone="B93000711"
 /tissue_type="cerebellum"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="10 days neonate"

misc_feature 1..1335
 /note="unknown: EST (GB|01805008, evidence: BLASTN, 99%, match=465);"

BASE COUNT 396 a 302 C 293 G 344 T

ORIGIN

Query Match 12.9%; Score 64.8; DB 11; Length 1335;
 Best Local Similarity 57.1%; Pred. No. 0.55;
 Matches 137; Conservative 0; Mismatches 102; Indels 1; Gaps 1;

QY 159 ATTGCTAAATTACTTGCAGAGATGCCACACAAATTCCTCCATCCTCATATATATGCG 218
 |||||
 Db 497 ACTGTAGTACATTGTTGCAGAGATGCCACAAATAACTCTCCATCCTCATATATACAC 556
 |||||

QY 219 CCGTTGCAATGTGACCTTGCCTACTTCTCTATCATCAAGATGTGGAGCTTATTTCCCATATA 278
 |||||
 Db 557 CCGTTGCAATGTGACCTTACCTTTCTCCCATGAGGGGTGGCATATCTTTCCCTGACTTT 616
 |||||

QY 279 TTGCACATAGAGTTGGCCTTCTGACTTGTCTTGCACATGGGATGTAGTACAAATGACACTG 338
 |||||
 Db 617 TTGAATCATTGCCTATTTTGTGTTGCTTAGACCACTGTTAGTGGCCAAA-GGACGT 675
 |||||

QY 339 TGCACCTTTGGATTTTAGGTTTGGAGAGAACTTACACCTTCCATCCTCATCTCTTGGAA 398
 |||||
 Db 676 ATGTATAGTGGAGCTTAGACTTCAAGAGGACATGACGCTTTTGGCCTCACCCCTCTAAGGA 735
 |||||

RESULT 14
 CCI42611/C
 LOCUS
 DEFINITION NDL.37D22.SP6 Notre Dame Liverpool Aedes aegypti genomic clone
 NDL.37D22, genomic survey sequence.

ACCESSION CCI42611
VERSION CCI42611.1 GI:30011666

KEYWORDS GSS.

SOURCE Aedes aegypti (yellow fever mosquito)
ORGANISM Aedes aegypti
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes.
 1 (bases 1 to 865)
 Loftus,B., Shetty,J., Knudson,D. and Severson,D.
 BAC end sequencing of Aedes aegypti
 Unpublished

REFERENCE Other_GSSs: NDL.37D22.77
 Contact: Brendan Loftus
 Department of Eukaryotic Genomics
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-3543
 Fax: 301-838-0208
 Email: enta@tigr.org
 Library was provided by David Severson
 Seq primer: SP6
 Class: BAC ends.

FEATURES Location/Qualifiers
 source
 1..865
 /organism="Aedes aegypti"
 /mol_type="genomic DNA"
 /strain="Liverpool"
 /db_xref="taxon:7159"
 /clone="NDL.37D22"
 /clone_lib="Notre Dame Liverpool"
 /note="Vector: pECBAC1; Site 1: Hind III; The library was prepared from whole body tissue of newly hatched L1 larvae by David Severson at the University of Notre Dame and Hongbin Zhang"

BASE COUNT 234 a 187 C 192 G 252 T

ORIGIN

Query Match 12.7%; Score 63.8; DB 29; Length 865;
 Best Local Similarity 65.1%; Pred. No. 0.76;
 Matches 110; Conservative 0; Mismatches 57; Indels 2; Gaps 1;

QY 205 CCTCATATATATGCCCTTTGCAATGTGACTTTGCTACTTC--TCTATCAAGATGTGGAG 262
 |||||
 Db 824 CCCTGTATGCATCTCTTTGCAATATCACTTTTAAAGTTCCTACCATCAAGAGTGCAG 765
 |||||

QY 263 CTTATTTTCCCATATATTGCACTAGAGTTGGCTTCTGACTTGTCTTGCACAAATGGAATGT 322
DB 764 CCAATTTCTCTAATCTTGGAAATCTGGACTGGCTTGTGACTTGTCTTGGCCCAACAAATGT 705
QY 323 AGTACAAATGACACTGTGCAACTTTTGGATTTTAGGTTTCGAGAGAACTT 371
DB 704 ATTAGAAATGACAGTGGCCAGTTCTTGAGCCTAGGCCCTCAAGAGGAGTT 656

RESULT 15
BX116065/c
LOCUS
DEFINITION
BX116065 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE998E035152 ; IMAGE:2090090, mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo;
1 (bases 1 to 497)
Ebert,D., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M.,
Radecki,U., Schneider,D. and Korn,B.
Human UnigeneSet - RZPD
Unpublished
Contact: Ina Rofls
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD; IMAGE998E035152;
RZPD; IMAGE998E035152;
Human UnigeneSet - RZPD (RZPD;IB No.092)
http://www.rzpd.de/CloneCards/cgi-bin/showLib.pl.cgi?response?libNo=972 Contact: Ina Rofls
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 5, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de

This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13r, Primer sequence: TTTCACACAGGAACACAGCTATGAC.
Location/Qualifiers
1..497
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE998E035152 ; IMAGE:2090090"
/lab_host="DH10B"
/clone_lib="Soares_NFL_T_GBC_S1"
/note="Organ: pooled; Vector: pTZ19-pac (Pharmacia) with
a modified polylinker; Site 1: Not 1; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbH19W, testis NHT, and B-cell
NCI-CGAP-GCB1) were mixed, and ss circles were made in
vitro. Following RAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bernaldo."

BASE COUNT 154 a 98 c 120 g 125 t
ORIGIN

Query Match
Best Local Similarity 56.4%; Score 63.6; DB 13; Length 497;
Matches 159; Conservative 0; Mismatches 119; Indels 4; Gaps 2;

QY 174 TTGCAGAGATGGCCACACAAATTCCTCTATCTCATATATATATGCCCCCTTGCATGTGA 233
DB 468 TTATAGATAGCCAAATCAAGTCTCTATCTTATGTCCTTATGCCCCACTCTCTTACAAATGTA 409

QY 234 CTTTGCTACTTCTCTATGCAAGATGTGGAGCTTATTTCCCATATATTGCACTAGAGTTGG 293
DB 408 CTGCCCTTTTTCCTCC-ACCAAGAGGTAGGGTCTATTTTGCTACCCCTCAGTCTGNACTAG 350
QY 294 CTTTCTGACTTGTCTTTCGACAAATGGAAATGTAGTACAAATGACACTGTGCAACTTTTGGATT 353
DB 349 CCAATGACTTGTCTTTCGACCAATCCAGTAGAGCAGAAGAGATGCTGTGACTTCTTAGGCC 290
QY 354 TAGGTTTCGAGAGAACTTACACCTTCCACTCCACTCCACTCTCTTTGGAACCCAGATGCAAT 410
DB 289 TCAACTTCAAGAGGTGTTTCAGACTTCCGCTCTCGTTCTCTTTCAACACTTCTGCCATGAC 230
QY 411 GTAAAGAAGTCAGGGCTATCTCTGCTAGAGACATATGTCCGAG 452
DB 229 GTAAAGAAATCTCAGTCTGTGTCTTCTCTTCTAGCCATTTATTCAG 188

Search completed: October 24, 2003, 22:51:01
Job time : 1050.9 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

DN nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:43:05 ; Search time 1384.09 Seconds
(without alignments)
14808.112 Million cell updates/sec

Title: US-09-830-902-1_COPY_50000_50500

Perfect score: 501

Sequence: 1 ggttttgggtgagtttctta.....atcagagagttggatagcaa 501

Scoring table: IDENTITY_NUC

Gapop 10.0 ; Gapext 1.0

Searched: 2895711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

- 1: gb_ha:
- 2: gb_htg:
- 3: gb_in:
- 4: gb_em:
- 5: gb_ov:
- 6: gb_pat:
- 7: gb_ph:
- 8: gb_pl:
- 9: gb_pr:
- 10: gb_ro:
- 11: gb_sts:
- 12: gb_sy:
- 13: gb_un:
- 14: gb_vt:
- 15: em_ba:
- 16: em_fun:
- 17: em_hum:
- 18: em_in:
- 19: em_mu:
- 20: em_or:
- 21: em_ov:
- 22: em_pat:
- 23: em_ph:
- 24: em_pl:
- 25: em_ro:
- 26: em_sts:
- 27: em_un:
- 28: em_vt:
- 29: em_hg_hum:
- 30: em_hg_inv:
- 31: em_hg_other:
- 32: em_hg_mus:
- 33: em_hg_pln:
- 34: em_hg_rdt:
- 35: em_hg_rdm:
- 36: em_hg_vrt:
- 37: em_sy:
- 38: em_hgoc_hum:
- 39: em_hgoc_mus:
- 40: em_hgoc_other:
- 41: em_hgoc_other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	501	100.0	110000	6	AX093471	AX093471 Sequence
2	501	100.0	110000	9	HSA246003	AC246003 Homo sapi
3	501	100.0	162692	9	CNS01058	AL121658 BAC seque
4	501	100.0	185281	2	AC011232	AC011232 Homo sapi
5	445.6	88.9	173106	9	AC068985	AC068985 Homo sapi
6	445.6	88.9	349938	6	AX647882	AX647882 Sequence
7	445	88.8	196953	9	AC009623	AC009623 Homo sapi
8	442.4	88.3	132868	2	AC021323	AC021323 Homo sapi
9	442.4	88.3	153323	9	AC022598	AC022598 Homo sapi
10	442.4	88.3	158404	9	AC090094	AC090094 Homo sapi
11	442.4	88.3	161644	9	AC025133	AC025133 Homo sapi
12	442.4	88.3	169505	9	CNS01070	AL132987 Human chr
13	441.8	88.2	99966	9	AC080012	AC080012 Homo sapi
14	440.8	88.0	91875	9	AL512286	AL512286 Human DNA
15	440.8	88.0	93089	9	AY308744	AY308744 Homo sapi
16	440.8	88.0	110000	2	AL831785	AL831785 Homo sapi
17	440	87.8	187525	2	AL354678	AL354678 Homo sapi
18	439.2	87.7	64965	9	AC006962	AC006962 Homo sapi
19	439.2	87.7	98056	9	AC006061	AC006061 Homo sapi
20	439.2	87.7	136631	9	AC109808	AC109808 Homo sapi
21	439.2	87.7	152951	9	AC026371	AC026371 Homo sapi
22	439.2	87.7	165590	9	AL512895	AL512895 Human DNA
23	439.2	87.7	215819	9	AC140490	AC140490 Homo sapi
24	439.2	87.7	222860	6	AX647883	AX647883 Sequence
25	439.2	87.7	243887	2	AC024898	AC024898 Homo sapi
26	438.8	87.6	157348	2	AP002446	AP002446 Homo sapi
27	438.8	87.6	169600	2	AC015958	AC015958 Homo sapi
28	438.8	87.6	176224	2	AP000898	AP000898 Homo sapi
29	438.8	87.6	184440	9	AP000919	AP000919 Homo sapi
30	438.8	87.6	185179	9	AP000937	AP000937 Homo sapi
31	438.4	87.5	155837	9	AL445495	AL445495 Human DNA
32	438.4	87.5	349999	6	AX647271	AX647271 Sequence
33	438.2	87.5	86141	9	EX323845	EX323845 Human DNA
34	438.2	87.5	182627	2	AC024262	AC024262 Homo sapi
35	438.2	87.5	182776	9	AC091078	AC091078 Homo sapi
36	438.2	87.5	191605	9	AC027670	AC027670 Homo sapi
37	438.2	87.3	41770	9	AC127090	AC127090 Homo sapi
38	437.6	87.3	73175	9	AC111193	AC111193 Homo sapi
39	437.6	87.3	74085	9	AC110780	AC110780 Homo sapi
40	437.6	87.3	79884	9	AC109440	AC109440 Homo sapi
41	437.6	87.3	86588	9	AC010453	AC010453 Homo sapi
42	437.6	87.3	114420	9	AL158076	AL158076 Human DNA
43	437.6	87.3	139550	9	AC110062	AC110062 Homo sapi
44	437.6	87.3	145824	9	AC090667	AC090667 Homo sapi
45	437.6	87.3	145824	9	AC090667	AC090667 Homo sapi

ALIGNMENTS

RESULT 1:
AX093471
LOCUS AX093471 110000 bp DNA linear PAT 30-MAR-2001
DEFINITION Sequence 1 from Patent WO01-8198.
ACCESSION AX093471
VERSION AX093471.1 GI:13509911
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Weissenbach, J. and Hazan, J.
TITLE Cloning, expression and characterisation of the spg4 gene
responsible for the most frequent form of autosomal spastic


```
5'UTR      9932..10056
           /gene="Spast"
CDS        join(10057..10471,33719..33805,35749..35832,45023..45118,
           60864..61052,61928..62062,62345..62438,73174..73248,
           74634..74705,82789..82864,83103..83194,83335..83414,
           88130..88172,89562..89641,91163..91233,93444..93484,
           100600..100722)
           /gene="Spast"
           /codon_start=1
           /product="spastin protein"
           /protein_id="CA860208.1"
           /db_xref="GI:6273493"
           /translation="MNSPGGRGKKKGGGASNPVPRPPPPCLAPAPPAAGPAPPPES
           FHKSNLYYFSYPLFVGFALLRLVAFHLGLLEFVWLQRFSSRALMAAKRSSGGAAPAPASA
           SAPAPVPGGEAEVRVFKQAFVLSIALRIDDEKAGQKEAVEWYKKGLELEKGI
           AVVTGQGEQCARRLQAKVYNLWAKORLGLLEKQVPLPFSKSCDVYNDSTNL
           ACRNGHLOESGAVPKRKOPLTHTSNLSPREKVMKTSAGLSGHRAPSKGLSVVS
           GVKQSGPAPTTHKGTFKTNRTNKPSTPTTATRKKKDLKNFRVDSNLANDLWNEIVD
           NGTAVKPDIDTAGQDLAKCALOEIVLPSLRPELFTGLPAPAPGLLLEGPPGSKTNLA
           KAAAESNATFFNISAASLTSKIVGEGEKUVRALFAVARELQPSLIFIDEVDSLQCEK
           REGHDSARRLKTEFLIEFDGVQAGGDDRVLYNGATNPPQELDAVLRREFKSVYVSL
           FNEETRLLLNKKCKGSPUTCLELAQLARNTDVGSGSMALAKCAAGSTIRELKP
           EQVKMSASEMRNIRLSDPTESLKKIKRSVSPOTLEAYIRNKKQEGCTTV"
           10472..33718
           /gene="Spast"
           /number=1
           33719..33805
           /gene="Spast"
           /number=2
           33806..35748
           /gene="Spast"
           /number=2
           35749..35832
           /gene="Spast"
           /number=3
           35833..45022
           /gene="Spast"
           /number=2
           45023..45118
           /gene="Spast"
           /number=4
           45119..60863
           /gene="Spast"
           /number=4
           60864..61051
           /gene="Spast"
           /number=5
           61052..61927
           /gene="Spast"
           /number=5
           61928..62061
           /gene="Spast"
           /number=6
           62062..62344
           /gene="Spast"
           /number=6
           62345..62438
           /gene="Spast"
           /number=7
           62439..73173
           /gene="Spast"
           /number=7
           73174..73248
           /gene="Spast"
           /number=8
           73249..74633
           /gene="Spast"
           /number=8
           74634..74705
           /gene="Spast"
           /number=9
           74706..82789
           /gene="Spast"
```

```
           /number=9
           82789..82864
           /gene="Spast"
           /number=10
           82865..83102
           /gene="Spast"
           /number=10
           83103..83194
           /gene="Spast"
           /number=11
           83195..83334
           /gene="Spast"
           /number=11
           83335..83414
           /gene="Spast"
           /number=12
           83415..88129
           /gene="Spast"
           /number=12
           88130..88172
           /gene="Spast"
           /number=13
           88173..89561
           /gene="Spast"
           /number=13
           89562..89641
           /gene="Spast"
           /number=14
           89642..91162
           /gene="Spast"
           /number=14
           91163..91233
           /gene="Spast"
           /number=15
           91234..93443
           /gene="Spast"
           /number=15
           93444..93484
           /gene="Spast"
           /number=16
           93485..100599
           /gene="Spast"
           /number=16
           100600..102009
           /gene="Spast"
           /number=17
           100721..102009
           /gene="Spast"
           3'UTR      101973..101978
           /gene="Spast"
           polyA_signal
BASE COUNT 30622 a 21640 c 22817 g 34921 t
ORIGIN
Query Match: 100.0%; Score 501; DB 9; Length 110000;
Best Local Similarity 100.0%; Pred. No. 2.8e-118;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1  GGTTTGGGTGAGTTTCTTAATCCTGAGTTCCTAGTTTGATTGCACTGTGGCCTGAGAGAC 60
        |||||
DB      50000  GGTTTGGGTGAGTTTCTTAATCCTGAGTTCCTAGTTTGTGCACTGTGGCCTGAGAGAC 50059
QY      61  AGTTTGTGTGAATTTCTGTTCCTTTACATTTTGCTGAGGAGTGTCTTAGTTCCAACTATGT 120
        |||||
DB      50060  AGTTTGTGTGAATTTCTGTTCCTTTACATTTTGCTGAGGAGTGTCTTAGTTCCAACTATGT 50119
QY      121  GGTCAATTTTGGGAATAGGTGTGGTGTGGTGTGCTGAGAAGAATGTATATTCTGTGATTTGG 180
        |||||
DB      50120  GGTCAATTTTGGGAATAGGTGTGGTGTGGTGTGCTGAGAAGAATGTATATTCTGTGATTTGG 50179
QY      181  GGTTAGAGTTCTGTAGATGCTATTAGTCCACTTGGTGACAGACTGAGTTCAAGTTCCCT 240
        |||||
DB      50180  GGTTAGAGTTCTGTAGATGCTATTAGTCCACTTGGTGACAGACTGAGTTCAAGTTCCCT 50239
```

```
QY 241 GGAATCTGTCTTGTGATCTGTCTAATATTGACAGTGGGGTGTGAAGTCTCCAGTATTA 300
DB 50240 GGATCTGTCTTGTGATCTGTCTAATATTGACAGTGGGGTGTGAAGTCTCCAGTATTA 50299
QY 301 TTGTGTGGGAGCTAAAGTCTCTTTGTAGGCTCTCAGGAGCTTGTATGAATCTGGGTG 360
DB 50300 TTGTGTGGGAGCTAAAGTCTCTTTGTAGGCTCTCAGGAGCTTGTATGAATCTGGGTG 50359
QY 361 CTCCTGTATTGGTGATATATATTAGGATAGTATAGTCTCTTCTTTGTAATGATCCCT 420
DB 50360 CTCCTGTATTGGTGATATATATTAGGATAGTATAGTCTCTTCTTTGTAATGATCCCT 50419
QY 421 TTAGCAATATATGATGGGCTCTTCTTCTCTCTTTGATCTCTTTGTTGTTTAAAGTCTGTT 480
DB 50420 TTAGCAATATATGATGGGCTCTTCTTCTCTCTTTGATCTCTTTGTTGTTTAAAGTCTGTT 50479
QY 481 TATCAGAGAGTTGGATTGCAA 501
DB 50480 TATCAGAGAGTTGGATTGCAA 50500

RESULT 3
CNS01DS8 162692 bp DNA linear PRI 16-APR-2002
LOCUS BAC sequence from the SPG4 candidate region at 2p21-2p22 BAC 563N04
DEFINITION of library RPCI-11 from chromosome 2 of Homo sapiens (human).
ACCESSION AL121658
VERSION AL121658.4 GI:20218783
KEYWORDS SPG4 genomic DNA interval.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 162692)
AUTHORS Hazan, J., Fonknechten, N., Navel, D., Paternotte, C., Samson, D.,
Artiguenave, F., Davoine, C.S., Cruaud, C., Durz, A., Wincker, P.,
Brottier, P., Catolico, B., Barbe, V., Burgunder, J.M.,
Prud'Homme, J.F., Brice, A., Fontaine, B., Heilig, R. and
Weissenbach, J.
TITLE Spastin, a novel AAA protein, is altered in the most frequent form
of autosomal dominant spastic paraplegia
JOURNAL Nat. Genet. (1999); in press
REFERENCE 2 (bases 1 to 162692)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (18-APR-2002) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT On Apr 19, 2002 this sequence version replaced gi:20160242.
FEATURES
source
location:Qualifiers
    ..162692
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="2"
    /clone="563N04"
    /clone_11b="RPCI-11"

BASE COUNT 44981 a 32017 c 32887 g 52792 t 15 others
ORIGIN
Query Match 100.0%; Score 501; DB 9; Length 162692;
Best Local Similarity 100.0%; Pred. No. 2,7e-116;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGTTCGCTGAGTTTCTTAACTCTGAGTCTCTAGTTGATTCGACTGTGGCCCTGAGAGAC 60
DB 31259 GGTTCGCTGAGTTTCTTAACTCTGAGTCTCTAGTTGATTCGACTGTGGCCCTGAGAGAC 31318
QY 61 AGTTGTGTAAATTTCTTCTTTTACATTTGCTGAGGAGTGCTTAGTCCCACTATGT 120
DB 31319 AGTTGTGTAAATTTCTTCTTTTACATTTGCTGAGGAGTGCTTAGTCCCACTATGT 31378
QY 121 GGTCATTTTGGAAATAGGTGTGTGTGGTCTGAGAGAAATGATATATTCTGTTGATTTGG 180
```

```
DB 31379 GGTCATTTTGGAAATAGGTGTGTGTGGTCTGAGAGAAATGATATATTCTGTGATTTGG 31438
QY 181 GGTTCAGATTTCTGTAGATGTCTATATAGTCCACATCTGTCAGAGAGTGAGTTCACTTCT 240
DB 31439 GGTTCAGATTTCTGTAGATGTCTATATAGTCCACATCTGTCAGAGAGTGAGTTCACTTCT 31498
QY 241 GGATCTGTCTTGTGATCTGTCTAATATTGACAGTGGGGTGTGGAAGTCTCCCAGTATTA 300
DB 31499 GGATCTGTCTTGTGATCTGTCTAATATTGACAGTGGGGTGTGGAAGTCTCCCAGTATTA 31558
QY 301 TTGTGTGGGAGCTAAAGTCTCTTTGTAGGCTCTCAGGAGCTTGTATGAATCTGGGTG 360
DB 31559 TTGTGTGGGAGCTAAAGTCTCTTTGTAGGCTCTCAGGAGCTTGTATGAATCTGGGTG 31618
QY 361 CTCCTGTATTGGTGATATATATTAGGATAGTATAGTCTCTTCTTTGTAATGATCCCT 420
DB 31619 CTCCTGTATTGGTGATATATATTAGGATAGTATAGTCTCTTCTTTGTAATGATCCCT 31678
QY 421 TTAGCAATATATGATGGGCTCTTCTTCTCTCTTTGATCTCTTTGTTGTTTAAAGTCTGTT 480
DB 31679 TTAGCAATATATGATGGGCTCTTCTTCTCTCTTTGATCTCTTTGTTGTTTAAAGTCTGTT 31738
QY 481 TATCAGAGAGTTGGATTGCAA 501
DB 31739 TATCAGAGAGTTGGATTGCAA 31759

RESULT 4
AC011232 185281 bp DNA linear HTG 10-MAR-2001
LOCUS Homo sapiens chromosome 2 clone RP11-78E13, WORKING DRAFT SEQUENCE,
DEFINITION 7 unordered pieces.
ACCESSION AC011232
VERSION AC011232.7 GI:13270720
KEYWORDS HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 185281)
AUTHORS Waterston, R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 185281)
AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Mar 10, 2001 this sequence version replaced gi:9795611.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WJGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: RH0078E13
----- Summary Statistics -----
Sequencing vector: M13; 57%
Sequencing vector: plasmid; 40%
Chemistry: Dye-primer ET; 48% of reads
Chemistry: Dye-terminator Big Dye; 50% of reads
Assembly program: Parap; version 0.990319
Consensus quality: 182798 bases at least Q40
Consensus quality: 183538 bases at least Q30
Consensus quality: 184045 bases at least Q20
Insert size: 168000; agarose-fp
Insert size: 184681; sum-of-contigs
Quality coverage: 7.07 in Q20 bases; agarose-fp
Quality coverage: 7.33 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
```



```
REFERENCE 3 (bases 1 to 173106)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (31-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 173106)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT On Oct 31, 2001, this sequence version replaced gi:16418054.
INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
gc-help@bcm.tmc.edu;

CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
http://gc.bcm.tmc.edu:8088/quality/info/genbank-annotation.html.

FEATURES
source Location/Qualifiers
1..173106
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="3"
/clone="RP11-217E22"
misc_feature 1..2001
/note="overlaps bases 12942..131142 of clone AC073156"
/function="clone overlap"
repeat_region 99..1507
/rpt_family="MSTC"
repeat_region 2220..2342
/rpt_family="MIR"
repeat_region 2951..3613
/rpt_family="LIM56"
STS 4275..4612
/standard_name="24462"
repeat_region 4553..4589
/rpt_family="(TG)n"
repeat_region complement(4816..5031)
/rpt_family="MIR"
```

```
repeat_region 5240..5296
/rpt_family="(TA)n"
repeat_region 5393..5394
/rpt_family="(TAGA)n"
repeat_region 5527..5920
/rpt_family="MER93"
repeat_region 6402..6521
/rpt_family="MIR"
repeat_region 6827..7020
/rpt_family="MIR"
repeat_region 8733..8757
/rpt_family="AT-rich"
repeat_region complement(8842..9152)
/rpt_family="AluJo"
repeat_region 9165..9187
/rpt_family="AT-rich"
repeat_region 9767..9931
/rpt_family="MSTD"
repeat_region 10272..10426
/rpt_family="MIR"
repeat_region 12479..12628
/rpt_family="AT-rich"
repeat_region 12730..13185
/rpt_family="MER89"
repeat_region 14101..14151
/rpt_family="AT-rich"
repeat_region 14576..14626
/rpt_family="AT-rich"
repeat_region 14758..14951
/rpt_family="HAL1"
repeat_region 15200..15558
/rpt_family="HAL1"
repeat_region complement(15909..16144)
/rpt_family="MER8"
repeat_region 16260..16295
/rpt_family="A-rich"
repeat_region 16365..16395
/rpt_family="AT-rich"
repeat_region complement(17961..18026)
/rpt_family="L2"
repeat_region complement(18027..18332)
/rpt_family="AluY"
repeat_region complement(18333..193346)
/rpt_family="L2"
repeat_region complement(19180..19424)
/rpt_family="L2"
repeat_region complement(19425..19914)
/rpt_family="LIPAS"
repeat_region complement(19815..20200)
/rpt_family="L2"
repeat_region complement(20532..20938)
/rpt_family="MLT2E"
repeat_region 20959..21521
/rpt_family="HERV16"
repeat_region 21645..22064
/rpt_family="HERV16"
repeat_region 22356..22486
/rpt_family="JTR50"
repeat_region 22509..22554
/rpt_family="CT-rich"
repeat_region complement(22555..22857)
/rpt_family="AluX"
repeat_region 22858..22873
/rpt_family="CT-rich"
repeat_region 22965..23767
/rpt_family="HERV2"
repeat_region 24216..24609
/rpt_family="DTR16A"
```

Query Match 88.9%; Score 445.6; DB 9; Length 173106;
Best Local Similarity 93.4%; Pred. No. 4.3e-104;
Matches 482; Conservative 0; Mismatches 19; Indels 15; Gaps 1;

Collins, S., Collymore, A., Cooke, P., DeAvallano, K., Dewar, K.,
Diaz, C.S., Dodge, S., Faro, S., Ferreira, P., FitzHugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Hagos, B., Hearford, A., Horton, L., Hulre, W.,
Iliev, I., Johnson, R., Jones, C., Karatas, A., Levine, R., Liu, G.,
Lamazares, R., Landers, T., Lebeckzy, J., Marquis, N., Matthews, J.,
MacLean, C., Macdonald, P., Marquis, N., Matthews, J., McCarthy, M.,
McEwan, P., McKernan, K., McPheeters, R., Meldrum, J., Merens, J.,
Mihova, I., Mienga, V., Murphy, T., Naylor, C., Nguyen, C., North, C.,
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phukhang, P., Pierre, N., Pollara, V., Raymond, C.,
Retta, R., Rieback, X., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S.,
Severy, P., Sougnuez, C., Spencer, B., Stange-Thorann, N.,
Stojanovic, N., Strauss, N., Subramanian, A., Talaras, C., Tesfaye, S.,
Theodore, C., Travers, M., Travis, N., Triciglio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, C., Zembek, J., Zimmer, A. and Zody, X.

Direct Submission

Submitted (01-MAY-2001) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On May 1, 2001 this sequence version replaced g1:23:3793.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997).

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project information

Center project name: L2082

Center clone name: 219_J_21

FEATURES

source

Location/Qualifiers

1. 196953
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="9"
/map="8"

/clone="RP11-219J21"

/clone_lib="RPC1-11 Human Male BAC"

complement(1031..1514)

/rpt_family="MER66B"

1650..1897

/rpt_family="LTRATG(n)"

complement(1200..1207)

/rpt_family="LIVE"

2281..2311

/rpt_family="AT_rich"

complement(1253..12630)

/rpt_family="ELAM C"

complement(1279..13085)

/rpt_family="AluSx"

complement(3288..4069)

/rpt_family="LIMB8"

complement(4069..4366)

/rpt_family="AluSc"

complement(4367..4824)

/rpt_family="LIM58"

complement(5100..5217)

/rpt_family="AluSg/x"

5233..6066

/rpt_family="L2"

6151..6455

/rpt_family="MLT1"

6583..6694

/rpt_family="L2"

complement(6698..6840)

/rpt_family="LIMED"

complement(6876..7080)

/rpt_family="LIMED"

repeat_region
complement(1708..17377)
/rpt_family="AluY"
repeat_region
complement(17378..17459)
/rpt_family="LIMED"
repeat_region
complement(17581..17786)
/rpt_family="LIMED"
repeat_region
complement(18027..18143)
/rpt_family="L1"
8434..18458
/rpt_family="AT_rich"
complement(18581..19162)
/rpt_family="LIMB3"
9277..19447
/rpt_family="MER5B"
complement(19498..19650)
/rpt_family="L1"
10365..10426
/rpt_family="CT-rich"
10614..10642
/rpt_family="AT_rich"
10643..10953
/rpt_family="AluSg1"
complement(10924..10931)
/note="Less than 30 qual SNGL region"
complement(11143..11438)
/rpt_family="AluJb"
11510..11614
/rpt_family="MIR"
11640..12135
/rpt_family="MLT2B2"
complement(12480..12919)
/rpt_family="MSTB"
13237..13353
/rpt_family="L2"
complement(15268..15576)
/rpt_family="AluSp"
complement(15742..15855)
/rpt_family="LIMA4A"
complement(15856..16163)
/rpt_family="AluY"
complement(16164..16923)
/rpt_family="LIMA4A"
complement(16924..17209)
/rpt_family="AluSx"
complement(17210..19516)
/rpt_family="LIMA4A"
complement(19517..19829)
/rpt_family="AluSg"
complement(19830..20438)
/rpt_family="LIMA4A"
complement(20439..20534)
/rpt_family="MER3C"
20535..20838
/rpt_family="AluSg"
complement(20839..20959)
/rpt_family="MER3C"
complement(20960..23361)
/rpt_family="LIMA4A"
complement(23384..23566)
/rpt_family="LIMA4A"
23694..23665
/rpt_family="L2"
23924..23965
/rpt_family="AT_rich"
complement(23976..24087)
/rpt_family="MER2"
complement(24093..30094)
/rpt_family="L1PA3"
24211..24223
/note="Less than 30 qual SNGL region"
complement(30095..30229)
/rpt_family="AluJb"
complement(30231..30324)
repeat_region

McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meid Grim, J.,
Meneus, L., Vinova, T., Mlenga, V., Murphy, T., Nayor, J., Nguyen, C.,
Norbu, C., Norman, C.H., O'Connor, T., O'Connell, P., O'Neill, G.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rowe, P.,
Roman, C., Rosetti, M., Roy, A., Santos, R., Schauer, S., Scrupback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, C., Tesfaye, S., Theodore, K.,
Topham, K., Travers, M., Travis, N., Trigg, C., J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, C., Ye, W. J., Young, S.,
Zairour, C., Zembek, L., Zimmer, A., and Zosy, M.

Direct Submission

Submitted (11-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Dec 11, 2001 this sequence version replaced g115421013.

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Dec 11, 2001 this sequence vers:cr replaced g:15421013.

All repeats were identified using Re

Smith, A. E. A. & Green, P. (1996-1997)

[illegible][illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Center: Whitehead Institute/MIT Center

Center code: W1BR

Web site: <http://www-seq.wi.mt>

Contact: sequence submission

Corec
normal

Salute

[illegible]

```
Location/Qualifiers
1. 25523
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
```

Query Match	88.3%	Score 442.4	DB 9	Length 15323
Best Local Similarity	93.0%	Pred. NO. 2.94-103		
Matches 480	Conservative	0	Mismatches 217	Indels 157 Gaps 1

QY	1	GGTTTGGGTGAGTTCTCTAACTCTGAGTCTAGTTGATTTGCACTCTGGCTGTGAGAGAC	60
DB			
QY	58112	GGTTTGGGTGAGTTCTCTAACTCTGAGTCTAGTTGATTTGCACTCTGGCTGTGAGAGAC	59053
DB			
QY	61	AGTTTGTGTAATTTCTGTCCTTTACATTTGCTGAGGAGTCTTTAGTTCCAACTATGT	120
DB			
QY	58052	AGTTTGTATAAATTTCTGTTCTTTTACATTTGCTGAGGAGTCTTTACTTCCAACTATGT	57993
DB			
QY	121	GGTCAATTTTGGAAATAGGTGTGGTGTGAGGAAGAAATGCTATATTTCTGTTGATTTGG	180
DB			
QY	57992	GGTCAATTTTGGAAATAGGTGTGGTGTGAGGAAGAAATGCTATATTTCTGTTGATTTGG	57933
DB			
QY	181	GGTTTAGAGATTCGTAGATGTCTATTAGGTCACACTTGGTGCAGAGACCTGAGTTCAGTTCCCT	240
DB			
QY	57932	GGTGGAGAGTTCTGTAGATGTCTATTAGGTCACACTTGGTGCAGAGACCTGAGTTCAGTTCCCT	57873
DB			
QY	241	GGATATCCCTTGTAAATTTTCTGCTCTGTTGATCTCTCTAATATTGACAGTGGGTGTTGA	285
DB			
QY	57872	GGATATCCCTTGTAAATTTTCTGCTCTGTTGATCTCTCTAATATTGACAGTGGGTGTTGA	57813
DB			
QY	286	AGTCTCCCAAGTATTATTGTGTGGGAGTCTAAGTCTCTCTTGTAGGTCCTCTAGGAGCTTGCT	345
DB			
QY	57812	AGTCTCCCAATTATTATTGTGTGGGAGTCTAAGTCTCTCTTGTAGGTCCTCTAAGGAGCTTGCT	57753
DB			
QY	346	TTATGAATCTGGTGTCTCTGTATTGGGTGCATATATATTAGGATAGTTAGCTCTTCTT	405
DB			
QY	57752	TTATGAATCTGGTGTCTCTGTATTGGGTGCATATATATTAGGATAGTTAGCTCTTCTT	57693
DB			
QY	406	GGTGAATTGATCCCTTTAGCATTTATGATGGCTCTCTTTCTCTTTTGATCTTTGTTG	465
DB			
QY	57692	GGTGAATTGATCCCTTTAGCATTTATGATGGCTCTCTTTCTCTTTTGATCTTTGTTG	57633
DB			
QY	466	GGTTAAAGTCTGTTTATCAGAGATTTGGATTGCCAA	501
DB			
QY	57632	GGTTAAAGTCTGTTTATCAGAGATTTGGATTGCCAA	57597
DB			

RESULT 10
ACC90094/C

```
LOCUS       AC090094               158404 bp      DNA      linear      PRI 15-MAR-2002
DEFINITION   Homo sapiens chromosome 8, clone RP11-252C19, complete sequence.
ACCESSION    AC090094
VERSION      AC090094.5   GI:115482342
KEYWORDS     HTG.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 158404)
    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  Homo sapiens chromosome 8, clone RP11-252C19
  Unpublished
  2 (bases 1 to 158404)
    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
    Barra,N., Bastien,V., Boguslavsky,L., Boukhalter,E., Brown,A.,
    Camarata,C., Campopiano,A., Choepel,Y., Colangelo,N., Collins,S.,
    Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
    Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
    Gardyna,S., Ginde,S., Goyette,M., Graham,J., Grand-Pierre,K.,
    Hagos,B., Heaford,A., Horton,L., Huime,W., Iliev,I., Johnson,R.,
    Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
    Lehoczy,J., Levine,R., Liu,G., Maclean,C., Macdonald,P.,
    Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
    McPheeters,R., Meldrim,J., Meneus,J., Mihova,T., Mienga,V.,
    Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
    O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
    Phunxhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
    Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,C., Rosetti,N.,
    Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
    Sougnès,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
    Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
    Travers,N., Travis,N., Triggilio,J., Vassiliev,H., Viel,R., Vo,A.,
    Wilson,B., Wu,X., Wyman,C., Ye,W.J., Young,G., Zainoun,J.,
    Zembek,J., Zimmer,A. and Zody,M.
  Direct Submission
  Submitted (14-FEB-2001) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  3 (bases 1 to 158404)
    Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
    Anderson,S., Barra,N., Bastien,V., Bloom,T., Boguslavsky,L.,
    Boukhalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,C.,
    Chazaro,B., Choepel,Y., Colangelo,N., Collins,S., Collamore,A.,
    Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
    Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
    Ginde,S., Gord,S., Goyette,M., Graham,J., Grand-Pierre,N.,
    Hagos,B., Horton,L., Huime,W., Iliev,I., Johnson,R., Jones,C.,
    Kamat,A., Karatas,A., Keills,C., LaRocque,K., Lamazares,R.,
    Landers,T., Lehoczy,J., Levine,P., Lindblad-Toh,K., Liu,G.,
    Maclean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
    McCarthy,M., McEwan,P., McKernan,K., McEwan,P., Meneus,J.,
    Mihova,T., Mienga,V., Murtuza,T., Naylor,J., Nguyen,C., Nicol,B.,
    Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
    Oliver,J., Peterson,K., Phunxhang,P., Pierre,N., Pollara,V.,
    Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
    Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
    Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
    Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
    Topham,K., Travers,M., Travis,N., Triggilio,J., Vassiliev,H.,
    Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
    Zainoun,J., Zembek,J., Zimmer,A. and Zody,M.
  Direct Submission
  Submitted (15-MAR-2002) Whitehead Institute/MIT Center for Genome
  Research, 320 Charles Street, Cambridge, MA 02141, USA
  On Mar 15, 2002 this sequence version replaced GI:14626341.
  All repeats were identified using RepeatMasker:
  Smit, A.P.A. & Green, P. (1996-1997)
  http://ftp.genome.washington.edu/RM/RepeatMasker.html
  ----- Genome Center
  Center: Whitehead Institute/ MIT Center for Genome Research
  Center code: WIBR
  Web site: http://www-seq.wi.mit.edu
  Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L12758
Center clone name: 252_C19

FEATURES             Location/Qualifiers
     source            ..158404
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-252C19"
                     /clone_lib="RPC111 Human Male BAC"
     repeat_region     733..564
                     /rpt_family="L2"
     repeat_region     complement(1037..1417)
                     /rpt_family="MLT2D"
     repeat_region     complement(1697..1825)
                     /rpt_family="LIME"
     repeat_region     1826..2120
                     /rpt_family="AluX"
     unsure            1960..1969
                     /note="<30 qual SNGL region"
     unsure            2000..2020
                     /note="<30 qual SNGL region"
     unsure            2030..2046
                     /note="<30 qual SNGL region"
     unsure            2086..2096
                     /note="<30 qual SNGL region"
     repeat_region     complement(2121..2899)
                     /rpt_family="LIME"
     unsure            2153..2157
                     /note="<30 qual SNGL region"
     repeat_region     2988..3249
                     /rpt_family="MLT2F"
     repeat_region     complement(5672..5745)
                     /rpt_family="MER5B"
     repeat_region     complement(7472..7524)
                     /rpt_family="L2"
     repeat_region     8965..9104
                     /rpt_family="MIR"
     repeat_region     10056..10077
                     /rpt_family="AT-rich"
     repeat_region     10154..10511
                     /rpt_family="L2"
     repeat_region     10770..10866
                     /rpt_family="L2"
     repeat_region     complement(11277..11584)
                     /rpt_family="AluX"
     repeat_region     complement(11650..11891)
                     /rpt_family="LIME1"
     repeat_region     complement(11994..12920)
                     /rpt_family="LIMEc"
     repeat_region     complement(12926..13195)
                     /rpt_family="AluSg"
     repeat_region     complement(13196..13833)
                     /rpt_family="LIMEc"
     repeat_region     complement(13820..15009)
                     /rpt_family="L1"
     repeat_region     complement(15010..15280)
                     /rpt_family="L2PA5"
     repeat_region     complement(15281..17128)
                     /rpt_family="L1"
     repeat_region     complement(17129..17260)
                     /rpt_family="FLAY_C"
     repeat_region     complement(17261..18749)
                     /rpt_family="L1"
     repeat_region     complement(18759..18838)
                     /rpt_family="HAL2"
     repeat_region     complement(18952..18923)
                     /rpt_family="LIMEc"
     repeat_region     complement(19633..19968)
                     /rpt_family="MER44D"
```


Lamazares, R., Landers, T., Jechoczky, J., Levine, R., Liu, G., MacLean, C., Macdonald, P., Major, J., Maquis, H., Matthews, C., McCarthy, X., McEwan, P., McKernan, K., McPeckers, P., Vaidyan, J., Veneus, L., Mihova, T., Mienga, V., Murphy, T., Nayler, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, C., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, A., Schauer, S., Scutback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thoman, N., Stojanovic, N., Strauss, K., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, E., Wu, X., Wyman, D., Ye, W., Young, S., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (01-DEC-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Nov 29, 2001 this sequence version replaced gu1604527.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WMR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: J7841
Center clone name: 227 F 6

FEATURES
SOURCE

Location/Qualifiers
: 162644

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"

```

```

/clone="RPI1-2236"
/clone : id="RPI1-11 : Human Male BAC"

```

complement: 1620. 1753.

```

/proc_family="VIR"
comp_jaton+ (0001 10001 10001

```

```

compres=1-004. . . 2363)
/zip family="MZR3"

```

complement (2309. 2500)

2022-09-27 14:27:27

```

case: "000",
  /? family="(?!)"

```

3640. 3812

3829-3878 Family = 50-100

```
/ret_family="Ga-vich"
```

3927. 4078

```

/IPC/amity=aga:icf:
complement(412: 4739)

```

```
/rpt_family="zyg9"
```

complement, 7938, .81431
/ref sam: v. "M-P"

complement: (8197, .8333)

```
/exp_family="AlusB"
```

9539. 4628
/ - Family - "TC" - "

20045, 19372

```
/xpt_family="(TAA);n"
```

repeat_region	/rpt_family="L1MA8"	complement(11689, .11697)
repeat_region	/rpt_family="L1MA8"	complement(11687, .12214)
repeat_region	/rpt_family="L1MA8"	12215, .12443
repeat_region	/rpt_family="L1PA8"	complement(12444, .13281)
repeat_region	/rpt_family="L1MA8"	13291, .13311
repeat_region	/rpt_family="AT-rich"	complement(13343, .13515)
repeat_region	/rpt_family="L2"	14013, .14089
repeat_region	/rpt_family="L1MA8"	15105, .16664
repeat_region	/rpt_family="L1MC2"	complement(16665, .16980)
repeat_region	/rpt_family="MER2"	16981, .18155
repeat_region	/rpt_family="L1MC2"	18172, .18899
repeat_region	/rpt_family="L1MC2"	19262, .19284
repeat_region	/rpt_family="AT-rich"	complement(19371, .21157)
repeat_region	/rpt_family="L1PA3"	21299, .21404
repeat_region	/rpt_family="L1MC4"	complement(21416, .21813)
repeat_region	/rpt_family="Alusq"	21869, .21908
repeat_region	/rpt_family="MIR3"	21918, .22013
repeat_region	/rpt_family="MIR3"	22015
unsure	/note="one of the reads c	
repeat_region	/rpt_family="L1MA8"	22562, .22589
repeat_region	/rpt_family="L1MA8"	complement(23312, .23612)
repeat_region	/rpt_family="Alusx"	24398, .24420
repeat_region	/rpt_family="CAAT:n"	complement(24805, .25228)
repeat_region	/rpt_family="MSTA"	complement(25253, .25943)
repeat_region	/rpt_family="MER31B"	complement(25972, .26127)
repeat_region	/rpt_family="MER31B"	complement(26155, .26267)
repeat_region	/rpt_family="L2"	complement(26495, .26619)
repeat_region	/rpt_family="MER20B"	complement(28926, .28931)
unsure	/note="30 qual SNGI regi	29086, .29269
unsure	/note="single clone cover	29131, .29141
repeat_region	/note="30 qual single cl	29203, .29498
repeat_region	/rpt_family="Alusq"	29712, .29751
repeat_region	/rpt_family="(TG)n"	complement(30588, .30705)
repeat_region	/rpt_family="L2"	complement(31556, .31748)
repeat_region	/rpt_family="MIR"	32920, .33038
repeat_region	/rpt_family="CT-rich"	34145, .34172
repeat_region	/rpt_family="AT-rich"	complement(34176, .34349)
repeat_region	/rpt_family="L1PB1"	


```
repeat_region complement(34350..34644)
repeat_region /rpt_family="AluY"
repeat_region complement(34645..35403)
repeat_region /rpt_family="L1P81"
repeat_region complement(36372..37043)
repeat_region /rpt_family="L1XC3"

Query Match 88.3%; Score 442.4; DB 9; Length 161644;
Best Local Similarity 93.0%; Pred. No. 2.9e-103;
Matches 480; Conservative 0; Mismatches 21; Indels 15; Gaps 1;

Y 1 GGTITTCGGTGAGTTTCTTAATCCTGAGTCTCTAGTTTCATTSCACATGCTGGCCCTGAGAGAC 60
Db 142079 GGTITTCGGTGAGTTTCTTAATCCTGAGTCTCTAGTTTCATTSCACATGCTGGCTGAGAGAC 142138

DY 61 AGTTTGTGTGTAATTTCTGTTCTTTTACATTTCCTGAGAGTCTCTTACATTCCTCAACTATGT 120
b 142139 AGTTTGTGTGTAATTTCTGTTCTTTTACATTTCCTGAGAGTCTCTGACTTCCTCAACTATGT 142196

Y 121 GGTCAATTTTGGAAATAGGTGTGGTGTGGTGGCTGAGAGAGATATATATCTCTGATTTGG 180
b 142199 GGTCAATTTTGGAAATAGGTGTGGTGTGGTGGCTGAGAGAGATATATATCTCTGATTTGG 142258

Y 181 GGTTCAGAGTCTGTAGATGTCTATTAGTCCACTTCCTGAGAGTCTGAGTTCAGTTTCCT 240
b 142259 GGTGGAGAGTCTGTAGATGTCTATTAGTCCACTTCCTGAGAGTCTGAGTTCAGTTTCCT 142318

Y 241 GGA-----TCTGCTTTTGTGATCTGTCTTAATATTTGACAGTGGGGTGTGA 285
b 142319 GGATATCCTTTGTTAACTTTCTGCTCTGATCTGTCTTAATATTTGACAGTGGGGTGTGA 142378

Y 286 AGTCTCCAGATATTATTGTGTGGAGTCTAAGTCTCTTTGTAGTCTCTAGGAGTCTGCT 345
b 142379 AGTCTCCAGATATTATTGTGTGTGGAGTCTAAGTCTCTTTGTAGTCTCTAGGAGTCTGCT 142438

Y 346 TTATGAATCTGGTGTCTCTGTATTTGGTGGCATATATATTTAGGATAGTTAGCTCTCT 405
b 142439 TTATGAATCTGGTGTCTCTGTATTTGGTGGCATATATATTTAGGATAGTTAGCTCTCT 142498

Y 406 GTTGAATGTATCCCTTTAGCATATATATGAGTCTCTTTGTCTCTTTGATCTCTTTGTTG 465
b 142499 GTTGAATGTATCCCTTTAGCATATATATGAGTCTCTTTGTCTCTTTGATCTCTTTGTTG 142558

Y 466 GTTTAAAGTCTGTCTTTATCAGAGAGTTCGATTGCAA 501
b 142559 GTTTAAAGTCTGTCTTTATCAGAGAGTTCGATTGCAA 142594

RESULT 12
NSCIDTV CNS01DTV 169505 bp DNA linear PRI 01-JUN-2003
DEFINITION Human chromosome 14 DNA sequence BAC R-188113 of library RPC1-11
from chromosome 14 of Homo sapiens (Human); complete sequence.
ACCESSION AL132987
VERSION 4.1 GI:14284836
KEYWORDS HTG; HTGS_ACTIVEFIN.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 169505)
AUTHORS Heilig,R., Petit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
Brottier,P., Cattolico,L., Barbe,V., Pelletier,E., Artiguenave,F.,
Levy,M., Eckenbrog,R., Bruls,T., deBerardinis,V., Cruaud,C.,
Gyapay,G., Saurin,W. and Weissbach,J.
Sequencing of the human chromosome 14
Unpublished
REFERENCE 2 (bases 1 to 169505)
Genoscope.
AUTHORS Direct Submission
TITLE Submitted (31-MAY-2003) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- web : www.genoscope.cns.fr)
```

```
COMMENT On Jun 3, 2001 this sequence version replaced g1:8217897.
----- Genoscope / Centre National de Sequencage
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr
-----
The following BAC sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : R-862G15 (AC=AL117192)
Downstream BAC (overlapping the SP6 end) : R-907G16 -----
Summary Statistics
Assembly program: Phrap; version 2.0
Quality coverage: 8.35x in Q20 bases; sum-of-contigs
-----
Overall quality chart :
Range : bases
0 :
1 - 9 :
10 - 19 : 3
20 - 29 : 6
30 - 39 : 43
40 - 49 : 716
50 - 59 : 1710
60 - 69 : 5293
70 - 79 : 19513
80 - 89 : 59301
90 - 99 : 32920
-----
Percentage of bases with a quality value >= 40 : 99 %
-----
FEATURES
source
Location/Qualifiers
1..169505
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="14"
/clone="R-188113"
/clone_lib="RPC1-11"
45652..45764
/contig="matching EMSL:FC1540
RHdb:RH26889
dbSTS:STS6768
Identified using the e-PCR software (G. Schuler)
47188..47291
/contig="matching EXBL:G32424
RHdb:RH67113
dbSTS:STS47040
Identified using the e-PCR software (G. Schuler)
-----
BASE COUNT 47640 a 36003 c 37377 g 48485 t
ORIGIN
-----
Query Match 88.3%; Score 442.4; DB 9; Length 169505;
Best Local Similarity 93.0%; Pred. No. 2.9e-103;
Matches 480; Conservative 0; Mismatches 21; Indels 15; Gaps 1;

QY 1 GGTITTCGGTGAGTTTCTTAATCCTGAGTCTCTAGTTTCATTSCACATGCTGGCCCTGAGAGAC 60
Db 29793 GGTITTCGGTGAGTTTCTTAATCCTGAGTCTCTAGTTTCATTSCACATGCTGGCTGAGAGAC 29852

QY 61 AGTTTGTGTGTAATTTCTGTTCTTTTACATTTCCTGAGAGTCTCTTACATTCCTCAACTATGT 120
Db 29853 AGTTTGTGTGTAATTTCTGTTCTTTTACATTTCCTGAGAGTCTCTTACATTCCTCAACTATGT 29912

QY 121 GGTCAATTTTGGAAATAGGTGTGGTGTGGTGGCTGAGAGAGATATATATCTCTGATTTGG 180
Db 29913 GGTCAATTTTGGAAATAGGTGTGGTGTGGTGGCTGAGAGAGATATATATCTCTGATTTGG 29972

QY 181 GGTTCAGAGTCTGTAGATGTCTATTAGTCCACTTCCTGAGAGTCTGAGTTCAGTTTCCT 240
Db 29973 GGTTCAGAGTCTGTAGATGTCTATTAGTCCACTTCCTGAGAGTCTGAGTTCAGTTTCCT 30032

QY 241 GGA-----TCTGCTTTTGTGATCTGTCTTAATATTTGACAGTGGGGTGTGA 285
Db 29973 GGTTCAGAGTCTGTAGATGTCTATTAGTCCACTTCCTGAGAGTCTGAGTTCAGTTTCCT 30032
```

b	30033	GGATACTGTTAACTTTCGTCTCGTTGATCTGCTAATGTTGACAGTGCGGTGTAA	30099
y	286	AGTCTCCCAAGTATTATTGTGTGGGAGTCTTAAGTCTCTTTTAGGTCTCTCAGGACTTCT	345
b	30093	AGTCTCCCATATTATTATTGTGTGGGAGTCTCTTTTAGGTCTCTCAGGACTTCT	30150
y	346	TATGAATCTGGGTGCCTCTCTATTTGGTGCCATATATTTAGGATAGTCTTCTTCT	405
b	30153	TATGAATCTGGGTGCCTCTCTATTTGGTGCCATATATTTAGGATAGTCTTCTTCT	30212
y	406	GTTGAATGATCCCTTTTAGCATATATATGATGGCCCTCTCTTTTGTCTTTTGTCT	465
b	30213	GTTGAATGATCCCTTTTAGCATATATATGATGGCCCTCTCTTTTGTCTTTTGTCT	30272
y	466	GTTTAAAGTCTGTTTTATCAGAGATTGGATTGCCAA	501
b	30273	GTTTAAAGTCTGTTTTATCAGAGATTGGATTGCCAA	30308

AC080012 99966 bp DNA linear FRI 31-MAY-2002
 Homo sapiens 12 BAC Rpl1-714K16 (Roswell Park Cancer Institute
 Human BAC Library) complete sequence.
 AC080012
 ACC80012.20 GI:14290372
 HTG.
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 : (bases 1 to 99966)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Aliosman,F.R., Allen,C.,
 Aisbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbara,J.,
 Benton,J., Birnbaum,K., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bukey,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carlton,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
 Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., DeLgado,O.,
 Denn,A.J., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
 Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
 Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
 Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
 Hart,X., Haviak,P., Hawes,A., Hernandez,J., Hernandez,O.,
 Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Horst,F.,
 Howard,S., Huber,J., Hulyk,S., Hume,C., Ioshikhes,I., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Joilivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratochvic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,D.C.,
 Lewis,D., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,C., Liu,W.,
 Loughsegh,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Ma,C., Maheshwari,V., Mapua,P., Marondel,I., Martin,R.,
 Martindale,A., Martinez,E., Massey,E., Mathewy,E., McLeod,K.P.,
 Meador,M., Mei,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,
 Xiner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
 Morris,S., Moser,M., Neal,D., Nelson,D., Newton,C., Newton,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokoko,S.,
 Ogih,M., Okworu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
 Quiles,Y., Ren,Y., Rives,M., Rojas,A., Rotubokan,I., Rolfe,M.,
 Ruiz,S., Savory,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
 Shochartz,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A.,
 Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A.,
 Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Tedford,B.,
 Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
 Vinson,R., Wall,R., Wang,S., Ward-Yore,S., Warren,R.,
 Washington,C., Watlington,S., Williams,G., Williamson,A.,
 Wiczyski,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,C.,
 Zorrilla,S., Zucherlapati,R. and Gibbs,R.


```
2y 181 GGTATTAGAGTTCTGTAGATGCTATTAGGTCCACTTGGTGCAGAGCTGAGTTCAGTTCC 240
   |||||
db 23543 GGTGGAGAGTTCTGTAGATGCTATTAGGTCCGCTTGGTGCAGAGCTGAGTTCAGTTCC 23602
   |||||
2y 241 GGA-----TCGTCTTGTGATCTGCTATATATGACAGTGGGGTGTGA 285
   |||||
db 23603 GGATATCCTTGTTAACTTTCTGTCTCGTTGATCTGCTATATGTTGACAGTGGGGTGTAA 23662
   |||||
2y 286 ACTCTCCAGTATTATTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTTAGGGACTTGT 345
   |||||
db 23663 ACTCTCCCAATTATTATTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTTAAGGACTTGT 23722
   |||||
2y 346 TTATGAACCTGGTCTCTGTATTTGGGTGCATATATATTAGGATAGTATAGCTTCTT 405
   |||||
db 23723 TTATGAACCTGGTCTCTGTATTTGGGTGCATATATATTAGGATAGTATAGCTTCTT 23782
   |||||
2y 406 GTTGAATTGATCCCTTTAGCAATTATATGATGGGCTTCTTTGTCTCTTTGATCTTTGTG 465
   |||||
db 23783 GTTGAATTGATCCCTTTAGCAATTATGTAAGGCTTCTTTGTCTCTTTGATCTTTGTG 23842
   |||||
2y 466 GTTAAAGTCTGTTTATCAGAGAGTTGGATTGCAA 501
   |||||
db 23843 GTTAAAGTCTGTTTATCAGAGAGTAGGATTGCAA 23878
   |||||
```

Search completed: October 24, 2003, 21:06:12
Job time : 1390.29 secs

FT exon 35749..35832
FT /*tag= f
FT /number= "3"
FT intron 35833..45022
FT /*tag= g
FT exon 45023..45118
FT /*tag= h
FT intron 45119..60863
FT /*tag= i
FT exon 60864..61051
FT /*tag= j
FT intron 61052..61927
FT /*tag= k
FT exon 61928..62061
FT /*tag= l
FT intron 62062..62344
FT /*tag= m
FT exon 62345..62438
FT /*tag= n
FT intron 62439..73173
FT /*tag= o
FT exon 73174..73246
FT /*tag= p
FT intron 73249..74633
FT /*tag= q
FT exon 74634..74705
FT /*tag= r
FT intron 74706..82788
FT /*tag= s
FT exon 82789..82864
FT /*tag= t
FT intron 82865..83102
FT /*tag= u
FT exon 83103..83194
FT /*tag= v
FT intron 83195..83334
FT /*tag= w
FT exon 83335..83414
FT /*tag= x
FT intron 83415..88129
FT /*tag= y
FT exon 88130..88172
FT /number= "12"
FT intron 88173..89561
FT /*tag= aa
FT exon 89562..89641
FT /*tag= ab
FT intron 89642..91162
FT /*tag= ac
FT exon 91163..91233

FT /*tag= ad
FT /number= "15"
FT intron 91234..93443
FT /*tag= ae
FT exon 93444..93484
FT /*tag= af
FT intron 93485..100599
FT /*tag= ag
FT exon 100600..102009
FT /*tag= ah
FT /number= "17"
XX PR2799138 A11
XX C9-MAR-2001
XX C3-SEP-1999; 99FR-0011097
XX C3-SEP-1999; 99FR-0011097
XX (CNRS) CNRS CENT NAT RECH SCI
XX Weissenbach C, Hazan C
XX WP1; 2001-283966/30
XX
PT New human nucleic acid from the SPG4 gene, useful e.g. for diagnosis of
PT autosomal dominant familial spastic paraplegia and in drug screening
XX
PS Claim 2; Page 45-106; 145pp; French.
XX
CC The present sequence represents a human SPG4 gene. The SPG4 gene encodes
CC a spastin polypeptide. Mutations in the SPG4 gene are responsible for
CC autosomal dominant familial spastic paraplegia. SPG4 polynucleotides,
CC and their fragments, are used to screen DNA banks for sequences that
CC encode spastin (particularly sequences in other mammals, specifically
CC mice); to identify SPG4 mutations, or other genetic anomalies,
CC particularly for diagnosis of autosomal dominant familial spastic
CC paraplegia (PSP-AD); to identify promoters and other regulatory elements
CC of the SPG4 gene; for detection and amplification; for recombinant
CC production of spastin; and for diagnostic genotyping of PSP-AD.
XX
SQ Sequence 110000 BP; 30622 A; 21640 C; 22917 G; 34921 T; 0 other;
Query Match 100.0%; Score 501; DB 22; Length 110000;
Best Local Similarity 100.0%; Pred. NO. 2.8e-125;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CY 1 GGTTTGGGTGAGTTTCTTAATCTGAGTCTCTAGTTTGATTGGACTGTGGCCTGAGAGAC 60
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 50000 GGTTTGGGTGAGTTTCTTAATCTGAGTCTCTAGTTTGATTGGACTGTGGCCTGAGAGAC 50059
CY 61 AGTTTGTGTGTAATTCTGTTCTCTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 50060 AGTTTGTGTGTAATTCTGTTCTCTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 50119
CY 121 GGTCAAATTTTGGAAATAGGTGTGGTGGTGGTCTGCTGAGAGAGATGTATATTTCTGTGATTTGG 180
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 50120 GGTCAAATTTTGGAAATAGGTGTGGTGGTGGTCTGCTGAGAGAGATGTATATTTCTGTGATTTGG 50179
CY 181 GGTTTAGAGTTCTGTAGATGCTCTATTAGTCCACTTGGTGCAGAGCTGAGTTCAAGTTCCCT 240
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 50180 GGTTTAGAGTTCTGTAGATGCTCTATTAGTCCACTTGGTGCAGAGCTGAGTTCAAGTTCCCT 50239
CY 241 GGATCTGCTCTTGTGATCTGTCTTAATATTGACAGTGGGGTGTGAAGTCTCCAGATTATA 300
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Dd 50240 GGATCTGCTCTTGTGATCTGTCTTAATATTGACAGTGGGGTGTGAAGTCTCCAGATTATA 50299
CY 301 TTGTCTGGGAGTCTAAGTCTCTTTGTAGTCTCTAGGGACTTGTCTTTATGAATCTGGGTG 360
Dd ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Dh 50300 TTGTGTGGGAGTCTAAGTCTCTTTGTAGTCTCTAGGGACTTGGCTTTATGAATCTGGGTG 50359
Qy 361 CTCCTGTATTGGTGCATATATATTAGGATAGTCTCTCTTGTGTGAATGATCCCT 420
Db 50360 CTCCTGTATTGGTGCATATATATTAGGATAGTCTCTCTTGTGTGAATGATCCCT 50419
Qy 421 TTACGATTATATGATGGCTTCTTCTTGTCTCTTTGATCTTTGTGTTTAAAGTCTGTTT 480
Db 50420 TTACGATTATATGATGGCTTCTTCTTGTCTCTTTGATCTTTGTGTTTAAAGTCTGTTT 50479
Qy 481 TATCAGAGAGTTGGATTGCAA 501
Db 50480 TATCAGAGAGTTGGATTGCAA 50500

RESULT 2
AAC69111
ID AAC69111 standard; DNA; 5065 BP.
XX
AC AAC69111;
XX
DT 31-JAN-2001 (first entry)
XX
DE Human secreted protein gene 3 clone H9CXL13.
XX
KW Cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein; ss.
X
US Homo sapiens.
X
X
N WO200055371-A1.
X
D 21-SEP-2000.
X
F 16-MAR-2000; 2000WO-US06783.
X
R 18-MAR-1999; 99US-0125055.
X
A (HVA-) HUMAN GENOME SCI INC.
X
X Ruben SM, Ni J, Ebner R, Rosen CA, Shi Y, Bixse C, Florence K;
X Komatsculis G, Lafleur DW, Monte PA, Olsen HS, Young PE;
X WPI; 2000-594448/56.
X P-PSDB; AA538011.

New nucleic acid molecules encoding 27 human secreted proteins for
diagnosing, preventing, treating or ameliorating medical conditions and
used as food additives or preservatives -
X
S Claim 1; Page 383-384; 453pp; English.
X
C The invention relate to the isolation of genes AAC69084-C69119 encoding
C 27 human secreted proteins AAB37994-B38019. The genes can be used to
C generate fusion proteins by linking to the gene for the human
C immunoglobulin G Fc portion (AAC69075) for increasing the stability of
C the fusion protein as compared to the human protein only. The genes and
C proteins are useful for preventing, ameliorating or treating medical
C conditions, e.g. by protein or gene therapy. The genes are isolated
C from a range of human tissues disclosed in the specification. The
C nucleic acids, proteins, antibodies and (ant)agonists are useful in
C the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
C and ovarian cancer, and other cancers of the adrenal gland, bone, bone
C marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
C (b) immune disorders e.g. Addison's disease, allergies, autoimmune
C haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
C disease, multiple sclerosis, rheumatoid arthritis and ulcerative
C colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
C wound healing; (e) neurological diseases e.g. cerebral anoxia and
C epilepsy; and (f) infectious diseases such as viral, bacterial, fungal

CC and parasitic infections.
XX
SQ Sequence 5065 BP; 980 A; 948 C; 1046 G; 2017 T; 74 other;
Query Match 87.0%; Score 436; DB 21; Length 5065;
Best Local Similarity 91.2%; Pred. No. 4.4e-108;
Matches 473; Conservative 4; Mismatches 24; Indels 15; Gaps 1;
Qy 1 GGTTTGGGTGAGTTTCTTTAATCCCTGAGTCTCTAGTTGATTGCACTGTGGCTTGAGAGAC 60
Db 3355 GGTTTGGGTGAGTTTCTTTAATCCCTGAGTCTCTAGTTGATTGCACTGTGGCTTGAGAGAY 3414
Qy 61 AGTTTGGTTGTAATTCTCTTCTTTTACATTTTCTGAGGAGTGTCTTTAGTTCCTCAACTATGT 120
Db 3415 AGTTTGGTTGTAATTCTCTTCTTTTACATTTTCTGAGGAGAGCTTCTACTTCCAAATATGT 3474
Qy 121 GGTCGAATTTGGATAGTGTGGTGGTGGTCTGAGAGAGAAATGTATATTTCTGTGATTGG 180
Db 3475 GGTCGAATTTGGATAGTGTGGTGGTGGTGGTGAAGAAATGTATATTTCTGTGATTGG 3534
Qy 181 GGTTTGAATTTCTGTAGATGTCTTATAGTCTCCACTGGTGCAGAGGTGAGTTTCACTTCT 240
Db 3535 GGTTGAGAGTCTGTAGATGTCTTATAGTCTCCGTGGTGGTGCAGAGGTGAGTTTCACTTCT 3594
Qy 241 GG-----ATCTGCTTTGTTGATCTGTCTAATATTGACAGTGGGTGTTGA 285
Db 3595 GGGTATCCTGTTRACCTTCTGTCTGTGATCTGTCTAATGTTCACAGTGGGTGTTAA 3654
Qy 286 AGTCTCCAGTATTATTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGAGTTGCT 345
Db 3655 AGTCTCCAGTATTATTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGAGTTGCT 3714
Qy 346 TTATGAATCTGGTGTCTCTGTATTGGTGCATATATATTTAGGATAGTTAGTCTTCTT 405
Db 3715 TTATGAATCTGGTGTCTCTGTATTGGTGCATATATATTTAGGATAGTTAGTCTTCTT 3774
Qy 406 GTTGAATTCATCCCTTTAGCATTATATGATGAGTGGCTCTTTGTCTCTTTTGTATCTTTG 465
Db 3775 GTTGAATTCATCCCTTTAGCATTATATGATGAGTGGCTCTTTGTCTCTTTTGTATCTTTG 3834
Qy 466 GTTTAAAGTCTCTTTTATCAGAGAGATTGGATTGCAA 501
Db 3835 GTTTAAAGTCTCTTTTATCAGAGAGATTGGATTGCAA 3870

RESULT 3
AAF33219
ID AAF33219 standard; cDNA; 5045 BP.
XX
AC AAF33219;
XX
DT 23-MAR-2001 (first entry)
XX
DE Human secreted protein gene 7 SEQ ID NO:17.
XX
KW Human; secreted protein; diagnosis; immunomodulatory; antisclerotic;
KW dermatological; immunosuppressive; antiinflammatory; anti-HIV;
KW immunostimulant; cytostatic; cardiant; vascular; anti-angiogenic;
KW ophthalmological; neuroprotectant; neurotropic; anticonvulsant; vulnary;
KW antialzheimers; antiparkinsonian; antimicrobial; immune disorder;
KW multiple sclerosis; systemic lupus erythematosus; HIV; infection;
KW hyperproliferative disorder; cancer; Gaucher's disease; wound healing;
KW cardiovascular disease; Scimitar syndrome; Chaga's cardiomyopathy;
KW coronary arteriosclerosis; angiogenic disorder; diabetic retinopathy;
KW corneal graft neovascularisation; neurological disorder; regeneration;
KW Huntington's chorea; Alzheimer's disease; Parkinson's disease;
KW infectious disease; chemotaxis; ss.
XX
OS Homo sapiens.
XX
PN WO200076530-A1.
XX
PD 21-DEC-2000.


```
Matches 476; Conservative 0; Mismatches 25; Indels 15; Gaps 1;
QY      1  GGTTTGGGTGAGTTCTTTAATCTGTAGTTCCTAGTTTGATTTTCACTGTGGCTGAGAGAC 40
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      12249 GGTTTGGGTGAGTTCTTTAATCTGTAGTTCCTAGTTTGATTTTCACTGTGGCTGAGAGAC 12308
QY      61  AGTTTGTGTGAATTCCTGTTCTTTTACATTTGCTGAGGAGTGTCTTAGTTTCAACTATGT 120
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      12309 AGTTTGTGATAATTCCTGTTCTTTTACATTTGCTGAGGAGTGTCTTAGTTTCAACTATGT 12368
QY      121  GGTCAATTTTGGAAATAGGTGTGTGTGCTGCTGAGAGAGATGATATATCTGTGATTGG 180
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      12369 GGTGAGTTTGGAAATAGGATGGTGTGTGCTGCTGAGAGAGATGATATATCTGTGATTGG 12428
QY      181  GGTTAGAGTCTCTGATGATGCTATTAGGTCACCTTGGTGCAGAGCTGAGTTCACTTCT 240
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      12429 GGTGAGAGTCTCTGATGATGCTATTAGGTCACCTTGGTGCAGAGCTGAGTTCACTTCT 12488
QY      241  GGA-----TCTGTCTTGTGATCTGTCTAATATTGACAGTGGGCTGTGA 295
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      12489 GGAATCCTTGTAACTTTCTGTGATGTGGATCTGTCTAATATTGACAGTGGGCTGTGA 12548
QY      286  AGTCTCCCACTATTATTGTGTGGAGTCTTAAGTCTCTTTGTAGTCTCTTAGGAGCTTCT 345
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      12549 AGTCTCCCACTATTATTGTGTGGAGTCTTAAGTCTCTTTGTAGTCTCTTAGGAGCTTCT 12608
QY      346  TTATGAATCTGGTCTCCTGTAATGGTGCAATATATATTAGATAGTTAGTCTTCTT 405
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      12609 TTATGAATCTGGTCTCCTGTAATGGTGCAATATATATTAGATAGTTAGTCTTCTT 12668
QY      406  GTTGAATTGATCCCTTTAGCAATTATATGATGGCTCTCTTGTCTCTTTGATCTTTGTG 465
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      12669 GTTGAATTGATCCCTTTATCATTTATGTAATGGCTCTCTTGTCTCTTTGACCTTTGTG 12728
QY      466  GTTTAAAGTCTGTTTATCAGAGAGTTGGATTGCAA 501
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      12729 GTTTAAAGTCTGTTTATCAGAGAGTGGATTGCAA 12764
```

RESULT 5

AAL35853

ID AAL35853 standard; DNA; 9192 BP.

XX

AC AAL35853;

XX

DT 08-JAN-2002 (first entry)

XX

DE Human musculoskeletal system related polynucleotide SEQ ID NO 2216.

XX

KW Cytostatic; immunosuppressive; neurotropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnery; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiac; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein;
KW musculoskeletal system; ds.

XX

CS Homo sapiens.

XX

PK WO200155367-A1.

XX

PD 02-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US01338.

XX

PR 31-JAN-2000; 2000US-0179065.

PR

PR 04-FEB-2000; 2000US-0180628.

PR

PR 24-FEB-2000; 2000US-0184664.

PR

PR 02-MAR-2000; 2000US-0186350.

PR

PR 16-MAR-2000; 2000US-0189874.

PR

PR 17-MAR-2000; 2000US-0190076.

PR

PR 18-APR-2000; 2000US-0198123.

PR

PR 19-MAY-2000; 2000US-0205515.

PR

PR 07-JUN-2000; 2000US-0209467.

PR

```
PR      28-JUN-2000; 2000US-0214886.
PR      30-JUN-2000; 2000US-0215135.
PR      07-JUL-2000; 2000US-0216647.
PR      07-JUL-2000; 2000US-0216880.
PR      11-JUL-2000; 2000US-0217487.
PR      11-JUL-2000; 2000US-0217496.
PR      14-JUL-2000; 2000US-0218290.
PR      26-JUL-2000; 2000US-0220963.
PR      26-JUL-2000; 2000US-0220964.
PR      14-AUG-2000; 2000US-0224518.
PR      14-AUG-2000; 2000US-0224519.
PR      14-AUG-2000; 2000US-0225213.
PR      14-AUG-2000; 2000US-0225214.
PR      14-AUG-2000; 2000US-0225266.
PR      14-AUG-2000; 2000US-0225267.
PR      14-AUG-2000; 2000US-0225268.
PR      14-AUG-2000; 2000US-0225270.
PR      14-AUG-2000; 2000US-0225447.
PR      14-AUG-2000; 2000US-0225757.
PR      14-AUG-2000; 2000US-0225758.
PR      14-AUG-2000; 2000US-0225759.
PR      18-AUG-2000; 2000US-0226279.
PR      22-AUG-2000; 2000US-0226681.
PR      22-AUG-2000; 2000US-0226868.
PR      22-AUG-2000; 2000US-0227182.
PR      23-AUG-2000; 2000US-0227309.
PR      30-AUG-2000; 2000US-0228324.
PR      01-SEP-2000; 2000US-0229287.
PR      01-SEP-2000; 2000US-0229343.
PR      01-SEP-2000; 2000US-0229344.
PR      01-SEP-2000; 2000US-0229345.
PR      01-SEP-2000; 2000US-0229509.
PR      05-SEP-2000; 2000US-0229513.
PR      05-SEP-2000; 2000US-0230437.
PR      05-SEP-2000; 2000US-0230438.
PR      08-SEP-2000; 2000US-0231242.
PR      08-SEP-2000; 2000US-0231243.
PR      08-SEP-2000; 2000US-0231244.
PR      08-SEP-2000; 2000US-0231413.
PR      08-SEP-2000; 2000US-0231414.
PR      08-SEP-2000; 2000US-0232080.
PR      08-SEP-2000; 2000US-0232081.
PR      12-SEP-2000; 2000US-0231968.
PR      14-SEP-2000; 2000US-0232197.
PR      14-SEP-2000; 2000US-0232398.
PR      14-SEP-2000; 2000US-0232399.
PR      14-SEP-2000; 2000US-0232400.
PR      14-SEP-2000; 2000US-0232401.
PR      14-SEP-2000; 2000US-0233063.
PR      14-SEP-2000; 2000US-0233064.
PR      14-SEP-2000; 2000US-0233065.
PR      21-SEP-2000; 2000US-0234223.
PR      21-SEP-2000; 2000US-0234274.
PR      25-SEP-2000; 2000US-0234997.
PR      25-SEP-2000; 2000US-0234998.
PR      26-SEP-2000; 2000US-0235484.
PR      27-SEP-2000; 2000US-0235834.
PR      27-SEP-2000; 2000US-0235835.
PR      29-SEP-2000; 2000US-0236327.
PR      29-SEP-2000; 2000US-0236367.
PR      29-SEP-2000; 2000US-0236368.
PR      29-SEP-2000; 2000US-0236369.
PR      29-SEP-2000; 2000US-0236370.
PR      02-OCT-2000; 2000US-0236802.
PR      02-OCT-2000; 2000US-0237037.
PR      02-OCT-2000; 2000US-0237039.
PR      02-OCT-2000; 2000US-0237039.
PR      13-OCT-2000; 2000US-0239935.
PR      13-OCT-2000; 2000US-0239937.
PR      20-OCT-2000; 2000US-0240960.
PR      20-OCT-2000; 2000US-0241221.
PR      20-OCT-2000; 2000US-0241785.
```


PN WO200155343-A1.
XX C2-AUG-2001.
PD
XX
PF 17-JAN-2001; 2001WO-US01322.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186150.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216890.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224513.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225256.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235936.

PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 02-OCT-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237017.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240560.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246414.
PR 08-NOV-2000; 2000US-0246415.
PR 08-NOV-2000; 2000US-0246416.
PR 08-NOV-2000; 2000US-0246417.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259676.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PR Rosen: CA, Barash SC, Ruben SM;


```
XX WPI; 2001-565190/63.
XX Nucleic acid encoding novel connective tissue associated polypeptides,
PT used in diagnosing, preventing, treating or ameliorating a disorder
PT such as cancer or rheumatoid arthritis.
XX Disclosure: SEQ ID No 1246; 673bp; English.
XX The present invention relates to the isolation of novel human connective
CC tissue related polypeptides (AAU86435-AAU86923) and the polynucleotide
CC (cDNA and genomic) sequences encoding them. The sequences of the
CC invention are useful in the diagnosis, treatment, prevention and/or
CC prognosis of diseases associated with connective tissue(s), including
CC cancer. The polynucleotide sequences of the invention are also useful
CC in gene therapy. ABK42:02-ABK43:16 represent genomic sequences encoding
CC the novel human connective tissue related polypeptides.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPQ
CC at ftp.wipo.int/pub/published_pat_sequences.
XX
SQ Sequence 9192 BP; 2180 A; 1723 C; 1845 G; 3444 T; 0 other;
Query Match 86.7%; Score 434.4; DB 23; Length 3192;
Best Local Similarity 92.1%; Pred. No. 146; 107;
Matches 475; Conservative 0; Mismatches 26; Indels 15; Gaps 1;
2Y 1 GGTCTTGGTGGAGTCTTAACTGAGTCTTCTAGTTCGATTCGACCTGGCTCGAGAGAC 60
2b 1 GGTCTTGGTGGAGTCTTAACTGAGTCTTCTAGTTCGATTCGACCTGGCTCGAGAGAT 6921
2Y 61 AGTTTGTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGGTTTGTTCCTCACTATGT 120
2b 61 AGTTTGTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGGTTTGTTCCTCACTATGT 6981
2Y 121 GGTCAATTTTGGATAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
2b 121 GGTCAATTTTGGATAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 7041
2Y 181 GGTCTAGAGTCTCTGATAGTCTCTATAGTCTCCACTGGTGGAGAGCTGAGTTCAGTTCCT 240
2b 181 GGTCTAGAGTCTCTGATAGTCTCTATAGTCTCCACTGGTGGAGAGCTGAGTTCAGTTCCT 7101
2Y 241 GGAATCTCCCTGTTAACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 285
2b 241 GGAATCTCCCTGTTAACTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7161
2Y 286 AGTCTCCAGATATTATTGTGGGAGTCTAAGTCTCTTTGTAGGCTCTCTAGGACTTGGT 345
2b 286 AGTCTCCAGATATTATTGTGGGAGTCTAAGTCTCTTTGTAGGCTCTCTAGGACTTGGT 7221
2Y 346 TTATGAATCTGGGTGCTCTGATTTGGTGCATATATATTAGGATAGTTCAGTTCCTTCT 405
2b 346 TTATGAATCTGGGTGCTCTGATTTGGTGCATATATATTAGGATAGTTCAGTTCCTTCT 7281
2Y 406 GTTGAATTCATCCCTTTAGCATATATCATGGCTCTTTTGTCTCTTTTGTCTTTTGTG 465
2b 406 GTTGAATTCATCCCTTTAGCATATATCATGGCTCTTTTGTCTCTTTTGTCTTTTGTG 7341
2Y 466 GTTAAAGTCTGTTTATCAGAGAGTTCGATTGCAA 501
2b 466 GTTAAAGTCTGTTTATCAGAGAGTTCGATTGCAA 7377
XX
D BX58841 standard; cDNA; 9192 BP.
X C BX58841;
X X
T 26-FEB-2003 (first entry)
X X
E cDNA encoding novel human musculoskeletal system antigen #1185.
```

```
XX
KW Gene; ss; musculoskeletal system antigen; Cancer; metastasis; re-vascularisation; thrombosis; arteriosclerosis; mineral content; cardiovascular condition; wound; injury; burn; angiogenesis; ulcer; post-operative tissue repair; limb regeneration; neuronal growth; neurodegenerative disorder; Alzheimer's disease; Parkinson's disease; AIDS-related complex; chondrocyte growth; bone regeneration; periodontal regeneration; tissue transport; bone graft; skin aging; keratinocyte growth; hair loss; melanocyte growth; cell proliferation; cell growth; organ transplant; cell differentiation; body height; weight; hair colour; eye colour; skin; percentage of adipose tissue; pigmentation; cosmetic surgery; metabolism; biorhythm; cardiac rhythm; depression; tendency for violence; pain; reproductive capability; hormone level; endocrine level; appetite; libido; memory; stress; storage capability; fat content; lipid content; protein content; carbohydrate content; vitamin content; cofactor content; nutritional component.
XX Homo sapiens.
OS
XX US2002147140-A1.
PN
XX 10-OCT-2002.
PL
XX 17-JAN-2001; 2001US-0764877.
XX
PR 31-JAN-2000; 2000US-179065P.
PR 04-FEB-2000; 2000US-180628P.
PR 28-JUN-2000; 2000US-214886P.
PR 07-JUL-2000; 2000US-216647P.
PR 07-JUL-2000; 2000US-216880P.
PR 11-JUL-2000; 2000US-217487P.
PR 11-JUL-2000; 2000US-217496P.
PR 14-JUL-2000; 2000US-218290P.
PR 26-JUL-2000; 2000US-220963P.
PR 26-JUL-2000; 2000US-220964P.
PR 14-AUG-2000; 2000US-224518P.
PR 14-AUG-2000; 2000US-224519P.
PR 14-AUG-2000; 2000US-225267P.
PR 14-AUG-2000; 2000US-225268P.
PR 14-AUG-2000; 2000US-225270P.
PR 14-AUG-2000; 2000US-225447P.
PR 14-AUG-2000; 2000US-225757P.
PR 14-AUG-2000; 2000US-225758P.
PR 22-AUG-2000; 2000US-226968P.
PR 30-AUG-2000; 2000US-228924P.
PR 01-SEP-2000; 2000US-229287P.
PR 01-SEP-2000; 2000US-229343P.
PR 01-SEP-2000; 2000US-229344P.
PR 01-SEP-2000; 2000US-229345P.
PR 05-SEP-2000; 2000US-229509P.
PR 05-SEP-2000; 2000US-229513P.
PR 08-SEP-2000; 2000US-231413P.
PR 21-SEP-2000; 2000US-234223P.
PR 21-SEP-2000; 2000US-234274P.
PR 25-SEP-2000; 2000US-234997P.
PR 27-SEP-2000; 2000US-235834P.
PR 29-SEP-2000; 2000US-236327P.
PR 29-SEP-2000; 2000US-236367P.
PR 29-SEP-2000; 2000US-236368P.
PR 29-SEP-2000; 2000US-236369P.
PR 29-SEP-2000; 2000US-236370P.
PR 02-OCT-2000; 2000US-236802P.
PR 02-OCT-2000; 2000US-237037P.
PR 02-OCT-2000; 2000US-237038P.
PR 02-OCT-2000; 2000US-237039P.
PR 02-OCT-2000; 2000US-237040P.
PR 13-OCT-2000; 2000US-239935P.
PR 20-OCT-2000; 2000US-240960P.
PR 20-OCT-2000; 2000US-241785P.
PR 20-OCT-2000; 2000US-241809P.
PR 01-NOV-2000; 2000US-244617P.
PR 17-NOV-2000; 2000US-249299P.
```


C producing transgenic animals and cells and also in gene therapy. This
C sequence represents a prostate specific nucleic acid described in
C the invention.

C Sequence 1885 BP; 662 A; 421 C; 395 G; 404 T; 0 other;

Query Match 86.1%; Score 432.8; DB 24; Length 1885;

Best Local Similarity 91.8%; Pred. No. 2.4e-107;

Matches 474; Conservative 0; Mismatches 27; Indels 15; Gaps 1;

Y 1 GGTGTTGGTGAGTTCTTAATCCTGAGTCTCTAGTTGATTCAGTCTGTGGCTGAGAGAC 60

b 733 GGTGTTGAGTGAGTTCTTAATCCTGAGTCTCTAGTTGATTCAGTCTGTGGCTGAGAGAC 674

Y 61 AGTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGAGTGCCTTAGTTCCAACTATGT 120

b 673 AGTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGAGTGCCTTAGTTCCAACTATGT 614

Y 121 GGTCAATTTTGAATAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180

b 613 GGTCAATTTTGAATAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 554

Y 181 GGTGTTAGAGTCTCTAGATGCTCTATTTAGTCTCCACTTGTGTCGAGAGTCTAGTTCCCT 240

b 553 GGTGGAGAGTCTCTAGATGCTCTATTTAGTCTCCACTTGTGTCGAGAGTCTAGTTCCCT 494

Y 241 GGA-----TCTGCTCTGTTGATCTGTCTAATATTGACAGTGGGGTCTTGA 285

b 493 GGAATATCCTTGTAACTTCTCTCTCATTTGATCTGTCTAATATTGACAGTGGGGTCTTGA 434

Y 286 AGTCTCCAGTATTATTGTGGGAGTCTAAGTCTCTTTTGTAGTCTCTAGGAGTCTTCT 345

b 433 AGTCTCCAGTATTATTGTGGGAGTCTAAGTCTCTTTTGTAGTCTCTAGGAGTCTTCT 374

Y 346 TTATGAATCTGGTCTCTCTCTATTTGGTGCATATATTTAGGATAGTCTAGTCTCTCT 405

b 373 TTATGAATCTGGTCTCTCTCTATTTGGTGCATATATTTAGGATAGTCTAGTCTCTCT 314

Y 406 GTTGAATTGATCTCTCTTTAGGATATATATGAGGCTCTTTTGTCTCTTTGATCTCTTTG 465

b 313 GTTGAATTGATCTCTCTTTAGGATATATATGAGGCTCTTTTGTCTCTTTGATCTCTTTG 254

Y 465 GTTGAAGTCTGTTTATACAGAGATTTGATTCGAA 50;

b 253 GTTGAAGTCTGTTTATACAGAGATTTGATTCGAA 218

AAZ35351 standard; DNA; 41599 BP.

AAZ35351;

27-MAR-2000 (first entry)

Cosmid including sequence spanning human chromosome 9p21.

Chromosome 9p21; human; cancer; tumour; diagnosis; prognosis;

methylthiodenosine phosphorylase; glioma; melanoma;

primary lymphoid malignancy; non-small cell lung cancer;

head and neck cancer; ovarian cancer; bladder cancer;

chondrosarcoma; ds.

Homo sapiens.

WO9967634-A1.

29-DEC-1999.

21-JUN-1999; 99WO-US13981.

23-JUN-1998; 98US-0690411.

17-JUN-1999; 99US-0335231.

(REGC) UNIV CALIFORNIA.

Carson DA, Schmid M, Carrera CC;

WPI; 2000-126550/11.

Diagnosing and determining prognosis for cancer causatively associated
with derangements of chromosome 9p21 -

Disclosure; Fig 7; 55pp; English.

This is the nucleotide sequence of a cosmid that includes sequences
spanning human chromosome 9p21 (ATCC AC00047). This region
harbours 5 genes with about 120 kb: the tumour suppressor genes
p16INK4A (p16) with its alternatively spliced form p10, p16INK4A
(p16) and p19ARF, and the gene for methylthiodenosine phosphorylase
(MTAP). The invention provides a method for diagnosis of, and
determining a prognosis for, cancer causatively associated with
derangements of chromosome 9p21. Underlying the invention is the
discovery that such derangements have their genes in deletions
occurring centromeric to STS 3.21, most often including breakpoints
in exon 8 and/or between exons 4 and 5 of the gene which encodes
MTAP. As the cancer and tumour development advance, deletions in
9p21 progress centromerically from the genesis point toward the
gene encoding p16. Thus, the method of the invention is performed
by determining whether (a) portions of the 9p21 region including
and telomeric to STS 3.21 are deleted, and (b) portions of the
9p21 region centromeric to STS 3.21 are deleted, where a positive
finding in step (a) and a negative finding in step (b) are
indicative of a cancer in an early stage of tumour development
and a positive finding in step (a) is indicative of a cancer in an
advanced stage of tumour development. Primer pairs (see AAZ35354-75)
are provided for use in claimed methods for diagnosing and
determining a prognosis for cancer associated with derangements of
9p21, especially a glioma, primary lymphoid malignancy, non-small
cell lung cancer, melanoma, head and neck cancer, ovarian cancer,
bladder cancer or a chondrosarcoma (claimed).

Sequence 41599 BP; 13477 A; 8216 C; 8242 G; 11664 T; 0 other;

Query Match 86.1%; Score 431.6; DB 21; Length 41599;

Best Local Similarity 91.8%; Pred. No. 1.3e-106;

Matches 472; Conservative 0; Mismatches 29; Indels 13; Gaps 1;

Y 1 GGTGTTGGTGAGTTCTTAATCCTGAGTCTCTAGTTGATTCAGTCTGTGGCTGAGAGAC 60

b 11896 GGTGTTGTTGAGTTCTTAATCCTGAGTCTCTAGTTGATTCAGTCTGTGGCTGAGAGAC 11837

Y 61 AGTGTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGTCTTAGTTCCAACTATGT 120

b 11834 AGTTTATATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGTCTTAGTTCCAACTATGT 11777

Y 121 GGTCAATTTTGGAAATAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180

b 11776 GGTCAATTTTGGAAATAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 11717

Y 181 GGTGTTAGAGTCTCTGAGATGTTCTATTAGTCCACTGGTGCAGAGCTGAGTTCCCT 240

b 11716 GGTGGAGAGTCTCTGAGATGTTCTATTAGTCTGCTGGTGGTGCAGAGCTGAATTCCT 11657

Y 241 GGA-----TCTGCTCTGTTGATCTGTCTAATATTGACAGTGGGTGTTGAAG 287

b 11656 GGATATCTCTTGTAACTTTGCTCATTTGATCTGTCTAATTTGTTAGTGGGTGTTAAAG 11597

Y 288 TCTCCAGTATTATTGTTGGGAGTCTAAGTCTCTTTTGTAGTCTCTAGGAGTCTGCTTT 347

b 11596 TCTCCAGTATTATTGTTGGGAGTCTAAGTCTCTTTTGTAGTCTCTAAGGAGTCTGCTTT 11537

Y 348 ATGAATCTGGTGTCTCTGTATTGGTGCATATATTTAGGATAGTCTAGTCTCTTCTGT 407

b 11536 ATGAATCTGAGTGTCTCTGTATTGGTGCATATATTTAGGATAGTCTAGTCTCTTCTGT 11477

QY 409 TGAATTGATCCCTTTAGCATATATGATGGGCTCTCTTTGCTCTCTTTGATCTTTGGT 467
|||||
11476 TGAATTGATCCCTTTACCATTAAGTAATGGGCTCTCTTTGCTCTCTTTGATCTTTGGT 11412
QY 463 TTAAAGTCTGTTTATCAGAGAGTTGGATTGCAA 501
|||||
11416 TTAAAGTCTGTTTATGAGAGACTAGGATTGCAA 11383

RESULT 10
AAS84982/c
D AAS84982 standard; cDNA; 2277 BP.
XX AC AAS84982;
DT 13-FEB-2002 (first entry)
DE DNA encoding novel human diagnostic protein #20786.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
DS WO200175067-A2.
XX 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2003; 2000US-0540217.
PPR 23-AUG-2003; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
DR WPI; 2001-639362/73.
DR P-PSDB; ABG20795.

New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity -
Claim 1: SEQ ID No 20786; 103pp; English.
The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. AAS64197-AAS94564 represent novel human
diagnostic coding sequences of the invention.
Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
Sequence 2277 BP; 927 A; 501 C; 411 G; 439 T; 0 other;

Query Match 86.1%; Score 431.2; DB 23; Length 2277;
Best Local Similarity 91.7%; Pred. No. 6.9e-107;
XX SQ

Matches 473; Conservative 0; Mismatches 28; Indels 15; Gaps 1;
QY 1 GGTTTGGGAGATTCTTAATCCTGAGTCTAGTTTGAATGCACCTGGGCTGAGAGAC 60
|||||
1123 GGTTTGGGAGATTCTTAATCCTGAGTCTAGTTTGAATGCACCTGGGCTGAGAGAC 1064
QY 61 AATTGTTGTAATTCCTGTTCTTTTACATTTGCTGAGGAGTCTTTAGTTCGAATATGT 120
|||||
1063 AATTGTTGTAATTCCTGTTCTTTTACATTTGCTGAGGAGTCTTTAGTTCGAATATGT 1004
QY 101 GGTCAATTTTGAATAGGTGGTGGTGGTGGTGGAGAGAAATGTATATTTGTTGATTTGG 180
|||||
1003 GGTCAATTTTGAATAGGTGGTGGTGGTGGTGGAGAGAAATGTATATTTGTTGATTTGG 944
QY 181 GGTTAGAGTTCTGTAGATGTTTATTAGTTCACCTTGGTGGAGAGCTGAGTTCAGTTTCT 240
|||||
943 GGTGGAGAGTTCTGTAGATGTTTATTAGTTCACCTTGGTGGAGAGCTGAGTTCAGATCTCT 884
QY 241 GGA-----TCTGCTTTGTTGATCTGTCTAATATTGACAGTGGGCTGTTGA 285
|||||
893 GGATATCCTTGTAACTTTCTGCTCTGTTGATCTGTCTAATATTGACAGTGGGCTGTTAA 824
QY 286 AGTCTCCAGTATTATTTGTTGGGAGTCTAAGTCTCTTTTGTAGGTTCTTAGGAGCTTGTCT 345
|||||
823 AGTCTCCAGTATTATTTGTTGGGAGTCTAAGTCTCTTTTGTAGGTTCTTAGGAGCTTGTCT 764
QY 346 TTATGATCTGGTCTGCTGCTGTTGTTGGTGGTGGATATATTTTAGGATAGTTAGTCTCTTT 425
|||||
763 TTATGATCTGGTCTGCTGCTGTTGTTGGTGGTGGATATATTTTAGGATAGTTAGTCTCTTT 704
QY 406 GTTGAATTGATCCCTTTAGCATATATGATGGGCTCTTTTGTCTCTTTTGTCTTTTGTGTTG 465
|||||
703 GTTGAATTGATCCCTTTAGCATATATGATGGGCTCTTTTGTCTCTTTTGTCTTTTGTGTTG 644
QY 466 GTTAAAGTCTGTTTATCAGAGAGTTTGGATTGCAA 501
|||||
643 GTTAAAGTCTGTTTATCAGAGAGTGGATTGCAA 608

RESULT 11
AAS75832/c
ID AAS75832 standard; cDNA; 2606 BP.
XX AC AAS75832;
DT 13-FEB-2002 (first entry)
XX DNA encoding novel human diagnostic protein #11636.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2003; 2000US-0540217.
PR 23-AUG-2003; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR P-PSDB; ABG21645.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess

SQ		Sequence 3316 BP; 1389 A; 713 C; 561 G; 653 T; 0 other;									
		Query Match 86.1%; Score 431.2; DB 23; Length 3316;									
		Best Local Similarity 91.7%; Pred. No. 7.8e-107;									
		Matches 473; Conservative 0; Mismatches 28; Indels 15; Gaps 1;									
QY	1	GGTTTGGGTGAGTTTCTTAATCCTGAGTTCAGTTCTAGTTTGCACTGGCCCTGAGAGAC	60								
DB	729	GGTTTGGGTGAGTTTCTTAATCCTGAGTTCGAGTTTGCACTGGCCCTGAGAGAC	670								
QY	61	AGTTTGGTGAATTTCTGTCTCTTTTACATTGCTGAGGAGTCTTTAGTCCCACTATCT	120								
DB	669	AGTTTGGTGAATTTCTGTCTCTTTTATATTGCTGAGGAGTCTTTAGTCCCACTATCT	610								
QY	121	GGTCAATTTGGGAATAGGTTGGTGTGGTCTGAGAAGAAATGATATATCTGTGATTTGG	180								
DB	609	GGTCAATTTGGGAATAAATGGTGTGTAGTCTGAGAAGAAATGATATATCTGTGATTTGG	550								
QY	181	GGTTTAGAGTCTGTAGATGTCTATATTAGTCCCACTTGGTCAGAGCTGAGTTCACTTCT	240								
DB	549	GGTGGAGAGTCTGTAGATGTCTATATTAGTCCCACTTGGTCAGAGCTGAGTTCAAAATCT	490								
QY	241	GGATATCCTTGTAACTTTCTGTCTCGTTGATCTGCTAAATGTTGACAGTGGGTTGTA	300								
DB	489	GGATATCCTTGTAACTTTCTGTCTCGTTGATCTGCTAAATGTTGACAGTGGGTTGTA	430								
QY	286	AGTCTCCCACTATTATTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTCT	345								
DB	429	AGTCTCCCACTATTATTGTGGGAGTCTAAGTCTCTTTGTAGGTCTCTAGGGACTTCT	370								
QY	346	TTATGAATCTGGTCTCTCTGTTATTGGTGGCATATATATTAGGATAGTCTCTCTT	405								
DB	369	TTATGAATCTGGTCTCTCTGTTATTGGTGGCATATATATTAGGATAGTCTCTCTT	310								
QY	406	GTTGAATTGATCCCTTTAGCAATATATGATGGCCCTCTTTGTCTCTTTGATCTTTGTG	465								
DB	309	GTTGAATTGATCCCTTTTACCATCATGTAATGGCCCTCTTTGTCTCTTTGATCTTTGTG	250								
QY	466	GTTTAAAGTCTGTTTATATCAGAGAGTTGGATTGCA	501								
DB	249	GTTTAAAGTCTGTTTATATCAGAGACTAGGATTGCA	214								

Search completed: October 24, 2003, 18:47:22
Job time : 131.251 secs

QY 240 TGG-----ATCTGCTTTGTTGATCTGTCTAATATTGACAGTGGGGTGTG 284
DB 19901 TGGGTATCCTTGTGTGACTTTTCTGTCTCGTTGATCTGTCTAATATTGACAGTGGGGTGTG 19960
QY 285 AAGTCTCCAGATTATTGTTGGGAGTCTAAGTCTCTTTGTAGGTCCTCTAGGACTTGC 344
DB 19961 AAGTCTCCAGATTATTGTTGGGAGTCTAAGTCTCTTTGTAGGTCCTCTAGGACTTGC 20020
QY 345 TTTATGAATCTGGGTCTCTCTCTATTTGGGTGCATATATATTAGGATAGTTAGTCTCTCT 404
DB 20021 TTTATGAATCTGGGTCTCTCTCTATTTGGGTGCATATATATTAGGATAGTTAGTCTCTCT 20080
QY 405 TGTGAATTGATCCCTTTAGCATATATGATGGGCTTCTTTGTCTCTTTTGTAGTCTTTGT 464
DB 20081 TGTGAATTGATCCCTTTAGCATATATGATGGGCTTCTTTGTCTCTTTTGTAGTCTTTGT 20140
QY 465 GGTAAAGTCTGTTTATCAGAGAGTGGATTGCAA 501
DB 20141 GGTAAAGTCTGTTTATCAGAGAGTGGATTGCAA 20177

RESULT 6
US-09-734-674-3
; Sequence 3, Application US/09734674
; Patent No. 6498022
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS.
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL031018
; CURRENT APPLICATION NUMBER: US/09/734,674
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 20201
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(20201)
; OTHER INFORMATION: n = A,T,C or G
US-09-734-674-3

Query Match 80.9%; Score 405.2; DB 4; Length 20201;
Best Local Similarity 89.9%; Pred. No. 3.5e-107;
Matches 464; Conservative 0; Mismatches 33; Indels 19; Gaps 2;
QY 1 GGTATTCGCTGAGTTCTTAATCCCTGAGTCTAGTTGATTCGACAGTGGGCTGAGAGAC 60
DB 145446 GGTATTCGCTGAGTCTTAATCCCTGAGTCTAGTTGATTCGACAGTGGGCTGAGAGAC 145505
QY 61 AGTTGTGTAAATTCCTGTTCTTTTACATTTGCTGAGGAGTGCTTACCTCAACTATGT 120
DB 145506 AGTTGTGTAAATTCCTGTTCTTTTACATTTGCTGAGGAGTGCTTACCTCAACTATGT 145565
QY 121 GGTCAATTTTGGAAATAGGTGTGGTGTGCTGAGAGAAATGTATATCTGTGATTTGG 180
DB 145566 GGTCAATTTTGGAAATAGGTGTGGTGTGCTGAGAGAAATGTATATCTGTGATTTGG 145625
QY 181 GGTATAGATTTCTGTAGATGCTATATAGCTCCACTGGTGCAGAGCTGAGTTCAGTCTCT 240
DB 145626 GGTATAGATTTCTGTAGATGCTATATAGCTCCACTGGTGCAGAGCTGAGTTCAGTCTCT 145685
QY 241 GG-----ATCTGCTTTGTTGATCTGTCTAATATTGACAGTGGGGTGTGGA 285
DB 145686 GGTATCCTTTGTCACACTTTCTGCTCTGTGATCTGTCTAATATTGACAGTGGGGTGTGGA 145745
QY 286 AGTCTCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCCTCTAGGACTTGCCT 345
DB 145746 AGTCTCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCCTCTAGGACTTGCCT 145805

QY 346 TTATGAATCTGGGTCTCTGTATTTGGGTGCATATATATTAGGATAGTTAGTCTCTCTT 405
DB 145806 TTATGAATCTGGGTCTCTGTATTTAGATACATATATATTAGGATAGTTAGTCTCTT 145865
QY 406 GTTGAATTGATCCCTTTAGCATATATGATGGGCTTCTTTGTCTCTTTTGTATCTTTG 465
DB 145866 GTTGAATTGATCCCTTTAGCATATATGATGGGCTTCTTTGTCTCTTTTGTATCTTTG 145921
QY 466 GTTAAAGTCTCTTTTATCAGAGAGTGGATTGCAA 501
DB 145922 GTTAAAGTCTCTTTTATCAGAGAGTGGATTGCAA 145957

RESULT 7
US-09-702-705-604
; Sequence 604, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Jodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Derrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 604
; LENGTH: 468
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(468)
; OTHER INFORMATION: n = A,T,C or G
US-09-702-705-604

Query Match 72.1%; Score 361.4; DB 4; Length 468;
Best Local Similarity 88.1%; Pred. No. 1.9e-95;
Matches 408; Conservative 0; Mismatches 40; Indels 15; Gaps 1;
QY 2 GTTTTGGGTGAGTTTCTTAATCCCTGAGTCTTAGTTGATTCGACACTGGCCCTGAGAGACA 61
DB 4 GTTTTGGGTGAGTTTCTTAATCCCTGAGTCTTAGTTGATTCGACACTGGCTGAGAGATA 63
QY 62 GTTTGTGTAAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCCAACTATGT 121
DB 64 GTTTGTGTAAATTTCTGTTCTTTTACACTTACTGAGGAGAGCTTTACTTCCAGATATGT 123
QY 122 GTCAATTTTGGAAATAGGTGTGGTGTGCTGAGAGAAATGTATATCTCTTTGATTTGGG 181
DB 124 GTGATTTTGGAAATAGGTGTGGTGTGCTGCTGAAAGAAATGTATATCTGTTGATTTGGG 183
QY 182 GTTAGAGTTCTGTAGATGCTATTAGGTCCTGCTGAGAGCTGAGTTCAGTTCCTCTG 241
DB 184 GTGAGAGTTCTGTANATGCTATTAGGTCCTGCTGAGAGCTGAGTTCAGTTCCTCTG 243
QY 242 GA-----TCTGCTCTTTGATCTGTCTAATATTGACAGTGGGGTGTGAA 286
DB 244 GATAGCCTTTGTTAACTTTCTGCTCTCTGTTGATCTGTCTAATATTGACAGTGGGGTGTGAA 303
QY 287 GTCTCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCCTCTAGGACTTGCCT 346
DB 304 GTCTCCAGTATTATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTCCTCTAAGGACTTGCCT 363

QY 347 TATGAATCTGGTGGTCTCTGTATGGTGGTGCATATATATTTAGGACAGTTAGCTCTCTTTG 406
|||||
Db 364 TATGAATCTGGTGGTCTCTGTATGGTGGTGCATATATTTAGGACAGTCTCTCTTTG 423
|||||
QY 407 TTGAATTGATCCCTTTAGCATATATGATGGCTCTCTTTGTC 449
|||||
Db 424 TTGAATTGATCCCTTTIACCATATATGTAATGGCTTGNCTCTTT 466
|||||

RESULT 8

US-09-736-457-604
Sequence 604, Application US/09736457
Patent No. 6509448
GENERAL INFORMATION:
APPLICANT: Wang, Tonglong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fargex, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darriek
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
APPLICANT: Wang, Aijun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 21021.478C15
CURRENT APPLICATION NUMBER: US/09/736,457
CURRENT FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 1864
SOFTWARE: FastSeq for Windows Version: 3.0
SEQ ID NO 604
LENGTH: 468
TYPE: DNA
ORGANISM: Homo sapien:
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(468)
OTHER INFORMATION: n = A,T,C or G

S-09-736-457-604

Query Match 72.1%; Score 361.4; DB 4; Length 468;
Best Local Similarity 88.1%; Pred. No. 1.9e-95;
Matches 408; Conservative 0; Mismatches 40; Indels 15; Gaps 1;

QY 2 GTTTGGTGAGTTCTTAAACCTGAGTTCTAGTTTGAATGACATGTGGCTGAGAGACA 61
|||||
Db 4 GTTTGAGTGAGTTCTTAAACCTGAGTTCTGAGTTGATGACATGTGGCTGAGAGACA 63
|||||
QY 62 GTTTGTTGTAATTCGTCTTTTACATTTGCTGAGGAGTGTATTAGTTCCAACTATGTG 121
|||||
Db 64 GTTTGTTATAATTCGTCTTTTACATTTACTGAGGAGAGTGTACTTCCAACTATGTG 123
|||||
QY 122 GTCAATTTTGGAAATAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 181
|||||
Db 124 GTCAATTTTGGAAATAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 183
|||||
QY 182 GTTCTAGAGTTCTGTAGATGTTTATAGTCCACTGGTGGAGAGCTGAGTTCAAGTCTG 241
|||||
Db 184 GTGGAGAGTTCTGTANATGTTCTATTAGTCCGTTGGTGGAGAGTTGAGTTCAATTCCTG 243
|||||
QY 242 GA-----TCGTCTGTGTGATCTGTCTAATATATGACAGTGGGTGTTGA 286
|||||
Db 244 GATAGCCTTGTTAACTTTCTGTCTGTGTGATCTGTCTAATATGACAGTGGGTGTTGA 303
|||||
QY 287 GTCTCCCAAGTATTATTGTTGGAGTCTAAGTCTCTTTGTAGCTCTCTAGGACTTGTCT 346
|||||
Db 304 GTCTCCCAATTATTATTGTTGGAGTCTAAGTCTCTTTGTAGTCACTAAGGACTTGTCT 363
|||||
QY 347 TATGAATCTGGTGGTCTCTGTATTGGTGGCATATATATTTAGGACAGTTAGCTCTCTTTG 406
|||||
Db 364 TATGAATCTGGTGGTCTCTGTATTGGTGGCATATATTTAGGACAGTCTCTCTTTG 423
|||||

QY 407 TTGAATTGATCCCTTTAGCATATATGATGGCTCTCTTTGTC 449
|||||
Db 424 TTGAATTGATCCCTTTIACCATATATGTAATGGCTTGNCTCTTT 466
|||||

RESULT 9

US-09-426-290-17c
Sequence 1, Application US/09426290
Patent No. 6410712
GENERAL INFORMATION:
APPLICANT: Berglind Ran Olafsdottir
APPLICANT: Jeffrey Gulcher
TITLE OF INVENTION: HDVAN NARCOLEPSY GENE
FILE REFERENCE: 2345.2001-000
CURRENT APPLICATION NUMBER: US/09/426,290
CURRENT FILING DATE: 1999-10-25
NUMBER OF SEQ ID NOS: 24
SOFTWARE: FastSeq for Windows Version: 4.0
SEQ ID NO 1
LENGTH: 168375
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (21181)...(21403)
NAME/KEY: CDS
LOCATION: (95252)...(95430)
NAME/KEY: CDS
LOCATION: (101753)...(101996)
NAME/KEY: CDS
LOCATION: (110324)...(110439)
NAME/KEY: CDS
LOCATION: (124058)...(124278)
NAME/KEY: CDS
LOCATION: (127009)...(127130)
NAME/KEY: CDS
LOCATION: (129910)...(129139)
US-09-426-290-1

Query Match 71.5%; Score 358; DB 4; Length 168375;
Best Local Similarity 84.1%; Pred. No. 1.5e-93;
Matches 434; Conservative 0; Mismatches 65; Indels 17; Gaps 2;

QY 1 GCTTTTGGTGAGTTTCTTAACTCTGAGTTCTAGTTTGGATGTCACCTGGCCCTGAGAGAC 60
|||||
Db 7781 GCTTTTGGTGAGTTTCTTAACTCTGAGTTCTAGTTTGGATGTCACCTGGCCCTGAGAGAC 7722
|||||
QY 61 AGTTTGTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGTCTTTAGTTCCAACTATGT 120
|||||
Db 7721 TGTGTTGTTATGATTTTAGTTCTTTGGCATTTGCTGAGGAGTGTCTTTACTTCCAATGATGT 7662
|||||
QY 121 GGTCAATTTTGGAAATAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
|||||
Db 7661 GGTCAATTTTGGAAATAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 7602
|||||
QY 181 GGTCTAGAGTTCTGTAGATGTCTATTAGTCCACTGGTGGAGAGCTGAGTTCAAGTCTCT 240
|||||
Db 7601 GGTGGAGAGTCTGTATAGATGTCTATTAGTCCACTGGTGGAGAGCTGAGTTCAAGTCTCT 7542
|||||
QY 241 GSA-----TCGTCTTTGTTGATCTGTCTAATATATGACAGTGGGTGTTGA 285
|||||
Db 7541 GAATGTACTTGTAAATTTCTGTCTCAATGATCTGTTTAAATATTGACAGTGGGTGTTAA 7482
|||||
QY 286 AGTCTCCCAAGTATTATTGTGGGAGTCTAAGTCTCTTTGTAGTCTCTAGGAGACTTGTCT 345
|||||
Db 7481 AGTCTCCCACTATTATTACGTGGGAGTCTAAGTCTC--TATAGTCTCTAAGAACTTGTCT 7424
|||||
QY 346 TTATGAATCTGGTGGTCTCTGTATTGGTGGCATATATATTTAGGATAGTCTCTCTTCT 405
|||||
Db 7423 TGATGAATTTGTGTGTCTCTGAAATGGGTGCATATATATTTAGGATAGTCTCTCTTCT 7364
|||||
QY 406 GTTGAATTGATCCCTTTAGCATTATATGATGGCCTTCTTTGTCTCTCTTTTGTCTTTGTTG 465
|||||

D0 7363 GTTGCAATGATCCCTTTATCATTAATGTAATGCCCTTTGTTGCTCTCTTTGATCTTTGTTG 7304
QY 466 GTTTAAAGTCTGTTTATACAGAGAGTTGGATTGCAA 501
D0 7303 GTTTGAGTCTATTTTATCAGAGAGACTAGGATTGCAA 7268

RESULT 10
US-09-385-982-157
; Sequence 157, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS: 11
; FILE REFERENCE: CCNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/417,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/099,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 157
; LENGTH: 668
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)....(668)
; OTHER INFORMATION: n = A,T,C or G

Query Match 54.4%; Score 272.6; DB 3; Length 666;
Best Local Similarity 86.2%; Pred. No. 1e-69;
Matches 374; Conservative 0; Mismatches 42; Indels 18; Gaps 6;
QY 1 GGTGTTGCGTGGATTTCTTAACTCCTGAGTTCTAGTTTGCATCTGCTGGCCTGAGAGAC 60
D0 53 GGTGTTGAGTGAGTTTCTTAACTGAGTTGCTGTTGATTGCTGCTGAGAGAC 112
QY 61 AGTTTGTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTCTTTAGTCCCACTAGT 120
D0 113 AGTTTGTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTCTTTAGTCCCACTAGT 172
QY 121 GGTCAATTTTGGAAATAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 180
D0 173 GGTCAATTTTGGAAATAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 232
QY 181 GGTGAGAGTCTGTAGATGCTATTAGTCCACTTGTGTCAGAGAGTGTGAGTTCAGTCTCT 240
D0 233 GGTGAGAGTCTGTAGATGCTATTAGTCTCTGCTGTCGAGAGAGTGTGAGTTCAGTCTCT 292
QY 241 G-----GATCTGCTTTGTTGATCTGTCTAAATTTGACAGTGGGGTGTGAG 287
D0 293 GATATCCTTGTGTAATTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 352
QY 288 TATCCCACTATTATTTGTTGAGTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 347
D0 353 TATCCCA-TATTATTTGTTGAGTCTTAA-TCTCTTTGATGTTCTTAAAGTCTTGT 410
QY 348 ATCAATCTGGTCTCTGTTATGGTGCATATATTTAGGATAGTTAGTCTCTCTCT 407
D0 411 AT-AACTGGTGTCTTG-ATGGGTGCA-ATATATTAGGATAGTTAGTCTCTCTCT 467
QY 408 TGAATGATCCCTT 421
D0 468 TGAATGGACCTTT 481

RESULT 11
US-09-318-448-11
; Sequence 11, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
; APPLICANT: Johnson, William G.
; APPLICANT: Stepien, Edward S.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318,448
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatencIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 18596
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-318-448-11
Query Match 45.9%; Score 230; DB 3; Length 18596;
Best Local Similarity 67.4%; Pred. No. 6.9e-57;
Matches 104; Conservative 0; Mismatches 25; Indels 19; Gaps 4;
QY 2 GTTTGCGTGGATTTCTTAACTCCTGAGTTCTAGTTTGCATCTGCTGGCCTGAGAGACA 61
D0 9059 GTTTGAGTGAGATTTCTTAACTCCTGAGTTCTAGTTTGCATCTGCTGGTGTGAGAGATA 9119
QY 62 GTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGTCTTACCTCACTATG 121
D0 9119 GTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGTCTTACCTCACTATG 9179
QY 122 GTCAATTTTGGAAATAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 181
D0 9179 GTGCGTTTGGAAATAGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 9238
QY 182 GTTTAGAGTCTCTGATGATGCTATTAGTCCACTTGTGTCAGAGTGTGAGTTCAGTCTCT 241
D0 9239 -ATGGAGTCTCTGATGATGCTATTAGTCTGCTGCTGATCTGTGACTGTGAGTTCAGTCTCT 9296
QY 242 G-----ATCTGCTCTGTTGATCTGTCTAATTTGACAGTGGGGTGTGAA 286
D0 9297 GGTATCCTTGTGAGTCTCTGCTGCTGATCTGTGACTGTGACTGTGAGTTCAGTTCCT 9355
QY 287 GTCTCCCACTATTATTGTGTGGGAGTCTAAGTCTCTTTTGTAGGTTCTCT 334
D0 9356 GTCTCCCACTATTATTGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 9402
RESULT 12
US-09-962-665-8
; Sequence 8, Application US/09962665
; Patent No. 6537759
; GENERAL INFORMATION:
; APPLICANT: Stanton, Jr., Vincent P.
; TITLE OF INVENTION: POLYGLUTAMATE SYNTHETASE GENE SEQUENCE
; TITLE OF INVENTION: VARIANCES HAVING UTILITY IN DETERMINING THE
; TITLE OF INVENTION: TREATMENT OF DISEASE
; FILE REFERENCE: 11926-015004
; CURRENT APPLICATION NUMBER: US/09/962,665
; CURRENT FILING DATE: 2001-09-24
; PRIOR APPLICATION NUMBER: 09/658,659
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 09/596,033
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 09/357,743
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 09/357,024
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: 60/093,484
; PRIOR FILING DATE: 1998-07-20
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0


```

; SEQ ID NO 8
; LENGTH: 18597
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 701, 1375;
; OTHER INFORMATION: n = c or a
; NAME/KEY: misc feature
; LOCATION: 716, 1293, 2401, 2429, 2618, 3083, 3125, 3635, 4256, 4898,
; LOCATION: 5062, 5167, 11069, 13229, 14479, 14730, 14796, 15344, 15450.
; LOCATION: 15533, 15590, 15840, 16149
; OTHER INFORMATION: n = a or g
; NAME/KEY: misc feature
; LOCATION: 732, 1379, 1590, 2488, 3212, 5006, 11238, 11422, 11686,
; LOCATION: 12598, 13171, 13645, 13782, 13856, 13813, 14586, 14788,
; LOCATION: 15042, 15546, 15770
; OTHER INFORMATION: n = c or t
; NAME/KEY: misc feature
; LOCATION: 1322, 1688
; OTHER INFORMATION: n = c or g
; NAME/KEY: misc feature
; LOCATION: 2594, 11293, 16199, 16203
; OTHER INFORMATION: n = g or t
; NAME/KEY: misc feature
; LOCATION: 3619
; OTHER INFORMATION: n = a or t
; NAME/KEY: misc feature
; LOCATION: 14547
; OTHER INFORMATION: nucleotide in position 14547 is t, or absent
;
; US-09-962-665-8

```

```

Query Match      45.9%; Score 230; DB 4; Length 18597;
Best Local Similarity 87.4%; Pred. No. 6.9e-57;
Matches 304; Conservative 0; Mismatches 25; Indels 19; Gaps 4;

Y 2 GTTTGGCTGAGTTCTTAATCCGAGTTCTAGTTGATTGACTGTGGCTGAGAGACA 61
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 9059 GTTTGAGTGAGATCTTAATCCGAGTTCTAGTTGATTGACTGTGGCTGAGAGACA 9113
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Y 62 GTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGTATTAGTTCCAACTATG 121
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 9119 GTTTGTTAATAATTTCTGTTCTTTTACATTTGCTGAGGAGTGTATTAGTTCCAACTATG 9178
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Y 122 GTCATATTTGGAATAGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 181
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 9179 GTGGGTTTGGAAATAGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG 9239
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Y 182 GTTTAGAGTTCTGTAGATGTCTATTAGTCCATTTGGTGTGGTGTGGTGTGGTGTGGTGTGG 241
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 9239 --ATGGAGTTCTGTAGATGTCTATTAGTCCATTTGGTGTGGTGTGGTGTGGTGTGGTGTGG 9296
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Y 242 G-----ATCTGCTCTTTGATCTCTATATATGACAGTGGGGTGTGAA 286
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 9297 GGTATCCTTGTGACTTTCTGTTCTGTTGATCTGTGCTACTGTGACAGT-GGGTGTAA 9355
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Y 287 GTCCTCCAGTATTATTTGTGGGAGTCTAAGTCTCTTTGAGGTTCT 334
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
b 9356 GTCCTCCAGTATTATTTGTGGT-GGAGTCTAAGTCTCTTTGAGGTTCT 3402
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 13
S-09-525-160B-9/c
Sequence 9, Application US/39525160B
Patent No. 6569681
GENERAL INFORMATION:
APPLICANT: Ivanov, Evguenii
TITLE OF INVENTION: METHODS OF IMPROVING HOMOLOGOUS RECOMBINATION
FILE REFERENCE: 10278/016001
CURRENT APPLICATION NUMBER: US/09/525,160B
CURRENT FILING DATE: 2000-03-14
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 9
; LENGTH: 3333
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-525-160B-9
;
; Query Match      38.4%; Score 192.2; DB 4; Length 3033;
; Best Local Similarity 65.5%; Pred. No. 3.1e-46;
; Matches 319; Conservative 0; Mismatches 153; Indels 15; Gaps 2;
;
QY 18 TTAATCCTGAGTTCTAGTTGATTGCACTGTGGCTGAGAGACAGTTGTTGTAATTTCT 77
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2074 TTATTGGGACTTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2015
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 78 GTTCTTTTACATTGCTGAGGAGTGTCTTAGTTCCAACTATGTGGTCAATTTTGGAAATAG 137
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2014 CTTTTTTCTAAATTTTAAGATTGTTTGTGACCTAGCATTTGGTATATCTCTTGAANT 1955
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 138 GTGCTGTGGTGTGCTGAGAGAAATGTAATTTCTGTTGATTTGGGGTTTGTAGTTCTCTAG 197
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1954 TATCCATGTGCTGAGGAGAGAAATGTGTATTTCTACAGCTGTGGAATCAAAATGTTCTGTAA 1895
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 198 ATGCTATATAGGTCCACTTTGGTGCAGAGCTGAGTTCA-----GTTCTCTGATCT 246
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1894 ATATCTACTAGGTCCATTGACCTATGGTGCAGATGAATTTGATGTTTCTTAATTTCT 1835
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 247 GTCTTGTGATCTGTCTAATATTGACAGTGGGGTGTGGAAGTCTCCAGTATTATTGTCT 306
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1834 GTCTGGATGATCTGTCCAAATGCTGAAAGTGGGGTGTGGAAGTCTCCAGTATTATTGTCT 1775
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 307 GGG----AGTCTAAGTCTCTTTGTAGTCTCTAGGGACTTGTCTTATGAATCTGGTCTCT 362
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1774 GGGGGTCTGTCTATCTCTCTCTTTGGCTCTAATATATTGCTTTTATATACCTGAGTACT 1715
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 363 CCTGTATTGGGTGCATATATATTAGGATAGTTAGTCTCTCTTGTGTAATTGATCCCTTT 422
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1714 CCAGTATTGGGTGAATATATATTAGAAATGTTAAATGCTCTTGTCTAAATGGACCCCTTT 1655
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 423 ACCATTATATGATGGCCTCTTTTGTCTCTTTTGTATCTTTGTTGTTTAAAGTCTGTTTAA 482
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1654 ATCACTATATATGACCTCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1595
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 483 TCAGAGA 489
|||
Db 1594 TCTGATA 1588
|||

```

```

RESULT 14
US-09-525-160B-4/c
Sequence 4, Application US/09525160B
Patent No. 6569681
GENERAL INFORMATION:
APPLICANT: Ivanov, Evguenii
TITLE OF INVENTION: METHODS OF IMPROVING HOMOLOGOUS RECOMBINATION
FILE REFERENCE: 10278/016001
CURRENT APPLICATION NUMBER: US/09/525,160B
CURRENT FILING DATE: 2000-03-14
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
;
; SEQ ID NO 4
; LENGTH: 3213
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-525-160B-4
;
; Query Match      38.4%; Score 192.2; DB 4; Length 3213;
; Best Local Similarity 65.5%; Pred. No. 3.1e-46;
; Matches 319; Conservative 0; Mismatches 153; Indels 15; Gaps 2;
;
QY 18 TTAATCCTGAGTTCTAGTTGATTGCACTGTGGCTGAGAGACAGTTGTTGTAATTTCT 77
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2074 TTATTGGGACTTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2015
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```
QY 78 GTCTTTTACATTTGCTGAGGAGTGTCTTAGTTCCAACTAATGGTCAATTTTGGAAATAG 137
Db 2014 CTTTCTTTTCTAAATTTAAGATTTGTTTGTGACCTAGCATTTGGTATATCCTTGAGAAAT 1955
QY 138 GTGTTGGTGTGGTGTGAGAAAGATGTATATCTGTTGATTTGGAGTTTAGAGTCTGTAG 197
Db 1954 TATCCATGTGCTGAGGAGAAAGATGTGTATTTCTACAGCTGTTGGATCAAATGTCTGTAA 1895
QY 198 ATGTCTATTAGGTCACCTTGGTGCAGAGCTGAGTTCA-----GTTCCTGGATCT 246
Db 1894 ATATCTACTAGGTCACCTTTGACCTATGGTGCAGATGAATTTGATGTTTGTAACTTTCT 1835
QY 247 GTCTGTTTGAATCTGTCTAAATATTGACAGATGGGGTGTGAAATGTCACAGTATTATTTGT 106
Db 1834 GTCTGGATGATCTGTCCAAATGCTGAAAGTGGGTGTGTAAGTCTCTTGTAAATGCTCTTGTAAATGGAGCCCTTT 1655
QY 307 GGG-----AGCTAAGTCTCTTTTGAAGTCTCTTAAGGACTTGGCTATGAATGTGGTGTCT 362
Db 1774 GGGGTCTGTCTATCTCTCTCTTGGCTCTTAATAATATTGGCTTATATATACCTGAGTACT 1715
QY 363 CCTGTATTGGTGCAATATATATTAGGATAGTTAGCTCTCTCTTGTGAATGTACCTTT 422
Db 1714 CCAGTATTGGTGAATATATATTAGAATGTTTAAATGCTCTTGTAAATGCTCTTGTAAATGGAGCCCTTT 1655
QY 423 AGCATATATATGAIGGCTCTCTTTGTCTCTTTTGAATCTTTGTGTTTAAAGTCTCTTTTA 482
Db 1654 ATCACTATATAATGAGCTCTCTCTCTCTCTTTTCAAGTCTTTGTGTTTGAATGTATTTTG 1595
QY 483 TCAGAGA 489
Db 1594 TCTGATA 1588

RESULT 15
US-09-220-132-168/c
; Sequence 168, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
; OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 07334-074001
; CURRENT APPLICATION NUMBER: US/09/220.132
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/068,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 168
; LENGTH: 9573
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-220-132-168

Query Match 37.4%; Score 187.4; DB 4; Length 9573;
Best Local Similarity 62.5%; Pred. No. 1.1e-44;
Matches 318; Conservative 0; Mismatches 176; Indels 15; Gaps 1;

QY 2 GTTTTGGTGTGAGTTCTCTTAATCCTGAGTCTTAGTTGATTTGCACCTGGCCCTGAGAGACA 61
Db 4074 GTTTTAAGGGTCTCTTTTGGAGTTAAATTTCCAGTTTATTCCCATTTGGTCTGAAAGGT 4015
QY 62 GTTTGTTGTAATTTCTCTTTTACATTTGCTGAGGAGTGTCTTTAGTTCCTCAACTATGTG 121
Db 4014 ACTTGACATAAATTTCAATTTCTTAAATTTGTTGAGACTTGTCTTGTGAGTCTGTTTGTGAGTCTATCTATG 3955
QY 122 GTCAATTTTGGAAATAGGTGTGGTGTGGTGTGAGAGATGTATATTCTGTTGATTGGG 181
Db 3954 ATTTATCTTGGAGAACGTTCTATGTGCTGATTAAATATGGTATATTCTGAAAGTTGTGG 3895
QY 182 GTTTAGAGTTCTGTAGATGTCTATTAGGTCCACTTTGGTGCAGAGCTGAGTTCAGTTTC 238
```

```
Db 3894 GTAGAAATGTTAGGTAAATATCTGCTAAATCTATTGTCTTAGCATATAATTTAAAGTCAAT 3835
QY 239 -----CTGGATCTGTCTTGTGTTGATCTGTCTTAATATTGACAGTGGGTGTTGAA 286
Db 3834 TTTTCTTTTATTGACTTTCTGCTTGTATGTCCTGTCTAGTGTCTCCATGSGAGTATTGAA 3775
QY 287 GTCTCCCAAGTATTATTGTGTGGAGTCTAAGTCTCTTTGTAGTCTCTCTTAGGACTTGTCTT 346
Db 3774 GTCCCTACTATTATTGTGTGGTSCCATCTATCCCATTTCTTAGGTCTAGTAATATTGTTT 3715
QY 347 TATGAATCTGGTGTCTCTGTATTGGGTSCCATATATATTTAGSATAGTTAGCTCTTCTTG 406
Db 3714 TATAAATTTGGGAGGTCCAGTATTAGGTSCATATATATTTAGSATTTGTATATTGTTCTG 3655
QY 407 TTGAATTGATCCCTTTTASCATTATAATGATGGCTTCTTTGTCTCTTTTGTATCTTTGTTGG 466
Db 3654 TTGGACCGATCCTTTTATCATTTATAAAATTTCTCTCTTTGTCTTTTATAACTGTGTTGC 3595
QY 467 TTTAAAGTCTCTTTTATCATCAGAGAGTTGGA 495
Db 3594 TTTAAAGCTTGTCTGATATATAAATAGGA 3566
```

Search completed: October 24, 2003, 15:42:57
Job time : 36.3542 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

DM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 15:28:54 : Search time 123.415 Seconds
(without alignments)
10886.662 Million cell updates/sec

Title: US-09-830-902-1_COPY_50000_50500

Perfect score: 501

Sequence: 1 ggttttgcgtgagttcttca.....atcagagagtgaggatgcaa 501

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1792395 seqs, 1340903451 residues

Total number of hits satisfying chosen parameters: 3534790

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCR_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCRUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	436	87.0	5065	14	US-10-050-882-39
2	436	97.0	1691139	14	US-10-067-534-1
3	434.4	86.7	9192	10	US-09-764-847-1246
4	434.4	86.7	9192	10	US-09-764-877-2218
5	434.4	86.7	9192	14	US-10-092-154-1246
6	431.2	86.1	5676	10	US-09-764-877-3756
7	431.2	86.1	32199	10	US-09-764-855-210
8	431.2	86.1	32199	14	US-10-072-349-210
9	431.2	86.1	250000	12	US-10-225-810-26
10	429.6	85.7	5446	11	US-09-764-891-7392
11	429.6	85.7	5671	10	US-09-764-877-3755
12	429.6	95.7	20746	11	US-09-764-891-10041
13	428	85.4	116592	10	US-09-818-512-3
14	428	85.4	116592	12	US-10-354-065-3
15	428	85.4	392000	12	US-10-027-983-11
16	428	85.4	465237	10	US-09-933-267A-1

17	428	85.4	786431	12	US-10-412-277-3	Sequence 3, Appli
18	426.4	85.1	3593	9	US-09-764-869-1634	Sequence 1634, Ap
19	426.4	85.1	3593	14	US-10-091-504-1634	Sequence 1634, Ap
20	426.4	85.1	6181	10	US-09-764-864-1674	Sequence 1674, Ap
21	426.4	85.1	6539	14	US-10-025-201-1	Sequence 1, Appli
22	426.4	85.1	6766	10	US-09-764-847-1878	Sequence 1878, Ap
23	426.4	85.1	6766	14	US-10-092-154-1878	Sequence 1878, Ap
24	425.4	84.3	168575	12	US-10-178-194-1	Sequence 1, Appli
25	424.8	84.8	4779	11	US-09-764-891-6770	Sequence 6770, Ap
26	424.8	84.8	4779	14	US-10-091-572-617	Sequence 617, App
27	424.8	84.8	7384	11	US-09-764-891-6771	Sequence 6771, Ap
28	424.8	84.8	7384	14	US-10-091-572-618	Sequence 618, App
29	424.8	84.8	30515	10	US-09-764-847-1208	Sequence 1208, Ap
30	424.8	84.8	30515	14	US-10-092-154-1208	Sequence 1208, Ap
31	424.8	84.8	32042	9	US-09-728-721-63	Sequence 63, Appli
32	424.8	84.8	32042	13	US-10-118-984-44	Sequence 44, Appli
33	424.8	84.8	32042	14	US-10-295-981-63	Sequence 63, Appli
34	424.8	84.8	34001	14	US-10-006-883A-15	Sequence 15, Appli
35	424.8	84.8	41104	9	US-09-816-685-3	Sequence 3, Appli
36	424.8	84.8	180557	13	US-10-003-806-6	Sequence 6, Appli
37	424.8	84.8	180557	13	US-10-003-806-9	Sequence 9, Appli
38	424	84.6	24132	9	US-09-764-860-661	Sequence 661, App
39	424	84.6	24132	14	US-10-074-095-661	Sequence 661, App
40	423.8	84.6	108359	14	US-10-191-807-3	Sequence 3, Appli
41	423.2	84.5	7380	11	US-09-764-891-6772	Sequence 6772, Ap
42	423.2	84.5	7380	14	US-10-091-572-619	Sequence 619, App
43	423.2	84.5	8041	11	US-09-764-891-6738	Sequence 6738, Ap
44	423.2	84.5	8041	11	US-09-764-891-6740	Sequence 6740, Ap
45	423.2	84.5	8041	14	US-10-091-572-588	Sequence 588, App

ALIGNMENTS

RESULT :

US-10-050-882-39
: Sequence 39, Application US/10050882
: Publication No. US2003010440CA
: GENERAL INFORMATION:
: APPLICANT: Ruben et al.
: TITLE OF INVENTION: 27 Human secreted proteins
: FILE REFERENCE: P2038P1
: CURRENT FILING DATE: 2002-01-18
: PRIOR APPLICATION NUMBER: 09/661,453
: PRIOR FILING DATE: 2000-03-13
: PRIOR APPLICATION NUMBER: PCT/US00/06783
: PRIOR FILING DATE: 2000-03-16
: PRIOR APPLICATION NUMBER: 60/125,355
: PRIOR FILING DATE: 1999-03-18
: NUMBER OF SEQ ID NOS: 156
: SOFTWARE: PatentIn Ver. 2.1.0
: SEQ ID NO 39
: LENGTH: 5065
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: 12531
: OTHER INFORMATION: n equals a,t,g, or c
: NAME/KEY: SITE
: LOCATION: 15063
: OTHER INFORMATION: n equals a,t,g, or c
US-10-050-882-39

Query Match 87.0%; Score 436; DB 14; Length 5065;

Best Local Similarity 91.7%; Pred. No. 1.3e+108;

Matches 473; Conservative 4; Mismatches 24; Indels 15; Gaps 1;

Oy : GCTTTGCGTGGTTCCTTAATCTGTAGTTCTGTAGTTGACTGTGGCTGAGAGAC 60

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 3355 GGTTTGAGTGAGGTTCTTAATCTGTAGTTCTGTAGTTGACTGTGGCTGAGAGAY 3414

```
QY 61 AGTTTGTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCTTTAGTTCCAACTATGT 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3415 AGTTTGTGTAATTTCTGTTCTTTTACATTTGCTGAGGAGAGCTTTACTTCCAACTATGT 3474
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 GGTCATATTTTGGAAATAGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3475 GGTCATATTTTGGAAATAGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3534
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 GGTTTACAGATTCTGTAGATGCTCTATAGTCCACTGGTGCAGAGCTGAGTTCACTTCTT 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3535 GGTTGAGAGATTCTGTAGATGCTCTATAGTCCACTGGTGCAGAGCTGAGTTCACTTCTT 3594
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 GG-----ATCTGCTCTGTGATCTGTCTAATATGACAGTGGGGTGTGGA 285
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3595 GGGTATCCCTTGTTTACTTTCTGTTGATCTGTCTAATATGACAGTGGGGTGTGGA 3654
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 286 AGTCTCCAGTATTATTATGTGGGAGTCTAAGTCTCTTTGAGGTTCTTAGGAGCTTGGT 345
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3655 AGTCTCCAGTATTATTATGTGGGAGTCTAAGTCTCTTTGAGGTTCTTAGGAGCTTGGT 3714
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 346 TTATGAAATCTGGGTGCTCTGTATTTGGGTGCATATATATTTAGGATAGTTAGCTTCTT 405
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3715 TTATGAAATCTGGGTGCTCTGTATTTGGGTGCATATATATTTAGGATAGTTAGCTTCTT 3774
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 406 GTTGAATTGATCCCTTTTACATTTATATGATGGCTCTTTTGTCTCTTTTGTATCTTTGG 465
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3775 GTTGAATTGATCCCTTTTACATTTATATGATGGCTCTTTTGTCTCTTTTGTATCTTTGG 3834
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 466 GTTAAAGTCTGTTTATCAGAGAGTGGATTGCAA 501
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3835 GTTAAAGTCTGTTTATCAGAGAGTGGATTGCAA 3870
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 2

```
US-10-067-514-1
; Sequence 1, Application US/10067514
; Publication No. US20030054531A1
; GENERAL INFORMATION:
; APPLICANT: Greta&sdottir, Solveig
; APPLICANT: Reynisdottir, Sigridur Th.
; TITLE OF INVENTION: HUMAN STROKE GENE
; FILE REFERENCE: 2345.2010-003
; CURRENT APPLICATION NUMBER: US/10/067,514
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 09/811/352
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1691139
; TYPE: DNA
; ORGANISM: Human
US-10-067-514-1
```

```
Query Match      87.0%; Score 436; DB 14; Length 1691139;
Best Local Similarity 92.2%; Pred. No. 1.6e-107;
Matches 476; Conservative 0; Mismatches 25; Indels 15; Gaps 1;
```

```
QY 1 GGTTTGGGAGATTCTTAATCCTGAGTTCTAGTTTATGATTCGACTGTGGCTTGAGAGAC 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12249 GGTTTGGAGTCTGTTTCTTAATCCTGAGTTCTAGTTTATGATTCGACTGTGGCTTGAGAGAC 12308
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 AGTTTGTGTAATTTCTGTTCTTTTACATTTTGGTCTGAGAGCTGCTTACTTCCAACTATGT 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12309 AGTTTGTGTAATTTCTGTTCTTTTACATTTTGGTCTGAGAGCTGCTTACTTCCAACTATGT 12308
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 GGTCATATTTTGGAAATAGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12369 GGTCAGTTTGGAAATAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 12428
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 GGTTTACAGATTCTGTAGATGCTCTATAGTCCACTGGTGCAGAGCTGAGTTCACTTCTT 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 12429 GGTGGAGAGTTCTGTAGATGCTCTATTAGTCCACTTGGTGCAGAGCTGAGTTCACTTCTT 12488
QY 241 GGA-----TCTGTCTTTGTTGATCTCTGCTAAATATTGACAGTGGGGTGTGGA 285
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12499 GGATATCCCTGTTAACTTTCTGTCATGTGGATCTGCTAAATGTTGACAGTGGGTGTGGA 12548
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 286 AGTCTCCAGTATTATTATGTGGGAGTCTAAGTCTCTTTGAGGTTCTTAGGGACTTGGT 345
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12549 AGTCTCCAGTATTATTATGTGGGAGTCTAAGTCTCTTTGAGGTTCTTAGGGACTTGGT 12608
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 346 TTATGAAATCTGGGTGCTCTGTATTTGGGTGCATATATATTTAGGATAGTTAGCTTCTT 405
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12609 TTATGAAATCTGGGTGCTCTGTATTTGGGTGCATATATATTTAGGATAGTTAGCTTCTT 12668
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 406 GTTGAATTCATCCCTTTAGCATTATATGATGGCTCTCTTTGCTCTTTTGTATCTTTGTTG 465
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12669 GTTGAATTCATCCCTTTATCAATTATGAAATGGCTCTCTTTGCTCTTTTGTATCTTTGTTG 12728
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 466 GTTAAAGTCTGTTTATCAGAGAGTGGATTGCAA 501
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 12729 GTTAAAGTCTGTTTATCAGAGAGTGGATTGCAA 12764
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 3

```
US-09-764-847-1246
; Sequence 1246, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P0009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1246
; LENGTH: 9192
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1246
```

```
Query Match      86.7%; Score 434.4; DB 10; Length 9192;
Best Local Similarity 92.1%; Pred. No. 4.6e-108;
Matches 475; Conservative 0; Mismatches 26; Indels 15; Gaps 1;
```

```
QY 1 GGTTTGGGAGTTTCTTAATCCTGAGTTCTAGTTTATGATTCGACTGTGGCTTGAGAGAC 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6862 GGTTTGGAGTGAATTC-TAACTCCTGAGTTCTAGTTTATGATTCGACTGTGGCTTGAGAGAT 6921
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 AGTTTGTGTAATTTCTGTTCTTTTACATTTGCTGAGAGTCTTACTTCCAAGTATGT 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6922 AGTTTGTGTAATTTCTGTTCTTTTACATTTGCTGAGAGAGCTTACTTCCAAGTATGT 6981
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 GGTCATATTTTGGAAATAGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6982 GGTCATATTTTGGAAATAGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 7041
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 GGTTTACAGATTCTGTAGATGCTCTATTAGTCCACTTGGTGCAGAGCTGAGTTCACTTCTT 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7042 GGTGGAGAGTTCTGTAGATGCTCTATTAGGTTCTGTTGTTGTCAGAGCTGAGTTCAACTTCTT 7101
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 GGA-----TCTGTCTTTGTTGATCTGCTAAATATTGACAGTGGGGTGTGGA 285
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7102 GGATATCCCTGTTAACTTTCTGTCCTGATCTGCTAAATGTTGACAGTGGGGTGTGTA 7161
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 286 AGTCTCCAGTATTATTATGTGGGAGTCTAAGTCTCTTTGAGGTTCTTAGGGACTTGGT 345
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7162 ATTCTCCCATATTATTATGTGGGAGTCTAAGTCTCTTTGAGGTTCTTAGGGACTTGGT 7221
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 346 TTATGAATCTGGGTGCTCTGTATTTGGGTGCATATATATTTAGGATAGTTAGCTTCTTCTT 405
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7222 TTATGAATCTGGGTGCTCTGTATTTGGGTGCATATATATTTAGGATAGTTAGCTTCTTCTT 7281
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

QY 406 GTTGAATTCATCCCTTAGCATATATATGATGCGCTTCTCTTGTCTCTTTTGTAGCTCTTGTG 465
DB 7282 GTTGAATTCATCCCTTTATCATATATGATGCGCTTCTCTTGTCTCTTTTGTAGCTCTTGTG 7341
QY 466 GTTAAAGTCCTGTTTATCAGAGAGATGGATTGCAA 501
DB 7342 GTTAAAGTCCTGTTTATCAGAGAGATGGATTGCAA 7377

RESULT 4
US-09-764-877-2218
; Sequence 2218, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 2218
; LENGTH: 9192
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-2218

Query Match 86.7%; Score 434.4; DB 10; Length 9192;
Best Local Similarity 92.1%; Pred. No. 4.6e-108;
Matches 475; Conservative 0; Mismatches 26; Indels 15; Gaps 1;
QY 1 GGTTCGGGTGAGTTCTTAAATCCCTAGTTTCTAGTTGATGCGCTTGTGCGCTGAGAGAC 60
DB 6862 GGTTCGGGTGAGTTCTTAAATCCCTAGTTTCTAGTTGATGCGCTTGTGCGCTGAGAGAT 6921
QY 61 AGTTTGTGTAATTCCTGCTCTTTTACATTTGCTGAGGAGTCTTTAGTCCCACTATGT 120
DB 6922 AGTTTGTGTAATTCCTGCTCTTTTACATTTGCTGAGGAGTCTTTAGTCCCACTATGT 6981

QY 121 GGTCAATTTTGGAAATAGGTGTGGTGTGCTGAGAGAGATGTAATTCCTGTTGATTTGG 180
DB 6982 GGTCAATTTTGGAAATAGGTGTGGTGTGCTGAGAGAGATGTAATTCCTGTTGATTTGG 7041
QY 181 GGTTCAGATTCCTGATGATCTCTATTAGTCCCACTGCTGAGAGAGTCTGATTCCTCT 240
DB 7042 GGTTCAGATTCCTGATGATCTCTATTAGTCCCTGCTGAGAGAGTCTGATTCCTCT 7101
QY 241 GGA-----TCTGCTCTTGTGATCTGCTAAATATTGACAGTGGGGTGTGA 285
DB 7102 GGATATCCCTGTAACTTCTGCTCTGCTGCTAAATATTGACAGTGGGGTGTGA 7161
QY 286 AGTCTCCCACTATTATTGTGGGAGCTAAGTCTCTTTGTAGTCTCTTAGGAGCTTCTCT 345
DB 7162 ATTCTCCCACTATTATTGTGGGAGCTAAGTCTCTTTGTAGTCTCTTAGGAGCTTCTCT 7221

QY 346 TTATGAATCTGGTCTCTCTGATTTAGGAGATATATTTAGGATAGTTAGCTCTCTCT 405
DB 7222 TTATGAATCTGGTCTCTCTGATTTAGGAGATATATTTAGGATAGTTAGCTCTCTCT 7281
QY 406 GTTGAATTCATCCCTTTAGCATATATGATGCGCTTCTCTTTGATCTTTGATCTTTGTTG 465
DB 7282 GTTGAATTCATCCCTTTATCATATGTAATGGCTTCTTTGTCTCTTTTGTATCTTTGTTG 7341
QY 466 GTTAAAGTCCTGTTTATCAGAGAGATGGATTGCAA 501
DB 7342 GTTAAAGTCCTGTTTATCAGAGAGATGGATTGCAA 7377

RESULT 5
US-10-092-154-1246
; Sequence 1246, Application US/10092154

; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior application removed - See File Wrapper or Palm
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 1246
; LENGTH: 9192
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1246

Query Match 86.7%; Score 434.4; DB 14; Length 9192;
Best Local Similarity 92.1%; Pred. No. 4.6e-108;
Matches 475; Conservative 0; Mismatches 26; Indels 15; Gaps 1;
QY 1 GGTTCGGGTGAGTTTCTTAAATCCCTAGTTTCTAGTTGATGCGCTTGTGCGCTGAGAGAC 60
DB 6862 GGTTCGGGTGAGATTCCTAAATCCCTAGTTTCTAGTTGATGCGCTTGTGCGCTGAGAGAT 6921
QY 61 AGTTTGTGTAATTCCTGCTCTTTTACATTTGCTGAGGAGTCTTTAGTCCCACTATGT 120
DB 6922 AGTTTGTGTAATTCCTGCTCTTTTACATTTGCTGAGGAGTCTTTAGTCCCACTATGT 6981
QY 121 GGTCAATTTTGGAAATAGGTGTGGTGTGCTGAGAGAGATGTAATTCCTGTTGATTTGG 180
DB 6982 GGTCAATTTTGGAAATAGGTGTGGTGTGCTGAGAGAGATGTAATTCCTGTTGATTTGG 7041
QY 181 GGTTCAGATTCCTGATGATCTCTATTAGTCCCACTGCTGAGAGAGTCTGATTCCTCT 240
DB 7042 GGTTCAGATTCCTGATGATCTCTATTAGTCCCTGCTGAGAGAGTCTGATTCCTCT 7101
QY 241 GGA-----TCTGCTCTTGTGATCTGCTAAATATTGACAGTGGGGTGTGA 285
DB 7102 GGATATCCCTGTAACTTCTGCTCTGCTGCTAAATATTGACAGTGGGGTGTGA 7161
QY 286 AGTCTCCCACTATTATTGTGGGAGCTAAGTCTCTTTGTAGTCTCTTAGGAGCTTCTCT 345
DB 7162 ATTCTCCCACTATTATTGTGGGAGCTAAGTCTCTTTGTAGTCTCTTAGGAGCTTCTCT 7221
QY 346 TTATGAATCTGGTCTCTCTGATTTAGGAGATATATTTAGGATAGTTAGCTCTCTCT 405
DB 7222 TTATGAATCTGGTCTCTCTGATTTAGGAGATATATTTAGGATAGTTAGCTCTCTCT 7281
QY 406 GTTGAATTCATCCCTTTAGCATATATGATGCGCTTCTCTTTGATCTTTGATCTTTGTTG 465
DB 7282 GTTGAATTCATCCCTTTATCATATGTAATGGCTTCTTTGTCTCTTTTGTATCTTTGTTG 7341
QY 466 GTTAAAGTCCTGTTTATCAGAGAGATGGATTGCAA 501
DB 7342 GTTAAAGTCCTGTTTATCAGAGAGATGGATTGCAA 7377

RESULT 6
US-09-764-877-3756
; Sequence 3756, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 3756
; LENGTH: 5676
; TYPE: DNA

```

; ORGANISM: Homo sapiens
US-09-764-877-3756

Query Match      86.1%; Score 431.2; DB 10; Length 5676;
Best Local Similarity 91.7%; Pred. No. 2.8e-107;
Matches 473; Conservative 0; Mismatches 28; Indels 15; Gaps 1;

QY 1 GGTTCGGGAGTTCCTTAACCTGAGTCTCTAGTTTGAATGAGTGGCCCTGAGAGAC 60
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 3323 GGTTCGGGAGTTCCTTAACCTGAGTCTCTAGTTTGAATGAGTGGCCCTGAGAGAC 3382

QY 61 AGTTTGTGTAATTTCTCTTTTACATTTTCTGAGGAGTCTCTAGTTTGAATGAGT 120
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 3383 AGTTTGTGTAATTTCTCTTTTACATTTTCTGAGGAGTCTCTAGTTTGAATGAGT 3442

QY 121 GGTCATTTTGGATAGGTGTGGTGGTCTGAGAGAAATGATATTTCTGTGATTTGG 180
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 3443 GGCCATTTTGGATAGGTGTGGTGGTCTGAGAGAAATGATATTTCTGTGATTTGG 3502

QY 181 GGTTTGTGATTTCTGTAGATGTCTATAGTCCACCTGGTGCAGAGTGGATTCCT 240
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 3503 GGTTTGTGATTTCTGTAGATGTCTATAGTCCACCTGGTGCAGAGTGGATTCCT 3562

QY 241 GGA-----TCGTCTTTGATCTGTCTAAATGACAGTGGGGTGTGA 285
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 3563 GGATATCTTTTAACTTCTGTCTAAATGACAGTGGGGTGTGA 3622

QY 286 AGTCTCCAGATTTATTTGTGGGAGTCTAAAGTCTCTTTGAGGCTCTAGGACTTCT 345
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 3623 AGTCTCCAGATTTATTTGTGGGAGTCTAAAGTCTCTTTGAGGCTCTAGGACTTCT 3682

QY 346 TTATGAATCTGGTGGTCTGTATTTGGTGCATATATTTAGGATAGTTCCTTCT 405
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 3683 TTATGAATCTGGTGGTCTGTATTTGGTGCATATATTTAGGATAGTTCCTTCT 3742

QY 406 GTTGAATGATCCCTTAGCATATATGATGAGTCTCTTTGATCTTTTGGTGG 465
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 3743 GTTGAATGATCCCTTAGCATATATGATGAGTCTCTTTGATCTTTTGGTGG 3802

QY 466 GTTAAAGTCTGTTTATCAGAGAGTTGGATTGCAA 501
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 3803 GTTAAAGTCTGTTTATCAGAGAGTTGGATTGCAA 3838

RESULT 7
US-09-764-855-210/c
; Sequence 210, Application US/09764855
; Patent No. US20020119919A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA10
; CURRENT APPLICATION NUMBER: US/09/764,855
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 210
; LENGTH: 32199
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-855-210

Query Match      86.1%; Score 431.2; DB 10; Length 32199;
Best Local Similarity 91.7%; Pred. No. 5.9e-107;
Matches 473; Conservative 0; Mismatches 28; Indels 15; Gaps 1;

QY 1 GGTTCGGGAGTTCCTTAACCTGAGTCTCTAGTTTGAATGAGTGGCCCTGAGAGAC 60
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 7319 GGTTCGGGAGTTCCTTAACCTGAGTCTCTAGTTTGAATGAGTGGCCCTGAGAGAC 7260

QY 61 AGTTTGTGTAATTTCTCTTTTACATTTTCTGAGGAGTCTCTAGTTTGAATGAGT 120
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 7319 AGTTTGTGTAATTTCTCTTTTACATTTTCTGAGGAGTCTCTAGTTTGAATGAGT 7260

QY 181 GGTTTGTGATTTCTGTAGATGTCTATAGTCCACCTGGTGCAGAGTGGATTCCT 240
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 7319 GGTTTGTGATTTCTGTAGATGTCTATAGTCCACCTGGTGCAGAGTGGATTCCT 7080

QY 241 GG-----ATCTGTCTTTGATCTCTAAATGACAGTGGGGTGTGA 285
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 7079 GGTATCTTTGTTGACTTTCTGTCTCTGTTGATCTCTAAATGTTGACAGTGGGGTGTGA 7020
```

```

DQ 7259 AGTTTGTGTAATTTCTCTTTTACATTTTCTGAGAGAGTCTCTTCAACTATGT 7200

QY 121 GGTCATTTTGGATAGGTGTGGTGGTCTGAGAGAAATGATATTTCTGTGATTTGG 180
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 7199 GGTCATATTTGGAGTAAAGTGTGGTGGTCTGAAAAGAAATGATATTTCTGTGATTTGG 7140

QY 181 GGTTTGTGATTTCTGTAGATGTCTATAGTCCACCTGGTGCAGAGTGGATTCCT 240
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 7139 GGTGGAGAGTTCCTGTAGATGTCTATAGTCCGCTGGTGCAGAGTGGATTCCT 7080

QY 241 GG-----ATCTGTCTTTGATCTCTAAATGACAGTGGGGTGTGA 285
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 7079 GGTATCTTTGTTGACTTTCTGTCTCTGTTGATCTCTAAATGTTGACAGTGGGGTGTGA 7020

QY 286 AGTCTCCAGATTTATTTGTGGGAGTCTAAAGTCTCTTTGAGGCTCTAGGACTTCT 345
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 7029 AGTCTCCAGATTTATTTGTGGGAGTCTAAAGTCTCTTTGAGGACTTCTAGGACTTCT 6960

QY 346 TTATGAATCTGGTGGTCTGTATTTGGTGCATATATTTAGGATAGTTCCTTCT 405
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 6959 TTATGAATCTGGTGGTCTGTATTTGGTGCATATATTTAGGATAGTTCCTTCT 6900

QY 406 GTTGAATGATCCCTTAGCATATATGATGAGTGGCTCTTTGATCTTTTGGTGG 465
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 6899 GTTGAATGATCCCTTAGCATATATGATGAGTGGCTCTTTGATCTTTTGGTGG 6840

QY 466 GTTAAAGTCTGTTTATCAGAGAGTTGGATTGCAA 501
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 6839 GTTAAAGTCTGTTTATCAGAGAGTTGGATTGCAA 6804

RESULT 8
US-10-072-349-210/c
; Sequence 210, Application US/10072349
; Publication No. US20030054430A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA1001
; CURRENT APPLICATION NUMBER: US/10/072,349
; CURRENT FILING DATE: 2002-02-11
; Prior application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: Patent In Ver. 3.1
; SEQ ID NO 210
; LENGTH: 32199
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-072-349-210

Query Match      86.1%; Score 431.2; DB 14; Length 32199;
Best Local Similarity 91.7%; Pred. No. 5.9e-107;
Matches 473; Conservative 0; Mismatches 28; Indels 15; Gaps 1;

QY 1 GGTTCGGGAGTTCCTTAACCTGAGTCTCTAGTTTGAATGAGTGGCCCTGAGAGAC 60
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 7319 GGTTCGGGAGTTCCTTAACCTGAGTCTCTAGTTTGAATGAGTGGCCCTGAGAGAC 7260

QY 61 AGTTTGTGTAATTTCTCTTTTACATTTTCTGAGGAGTCTCTAGTTTGAATGAGT 120
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 7259 AGTTTGTGTAATTTCTCTTTTACATTTTCTGAGGAGTCTCTAGTTTGAATGAGT 7200

QY 121 GGTCATTTTGGATAGGTGTGGTGGTCTGAGAGAAATGATATTTCTGTGATTTGG 180
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 7199 GGTCATATTTGGAGTAAAGTGTGGTGGTCTGAAAAGAAATGATATTTCTGTGATTTGG 7140

QY 181 GGTTTGTGATTTCTGTAGATGTCTATAGTCCACCTGGTGCAGAGTGGATTCCT 240
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 7139 GGTGGAGAGTTCCTGTAGATGTCTATAGTCCGCTGGTGCAGAGTGGATTCCT 7080

QY 241 GG-----ATCTGTCTTTGATCTCTAAATGACAGTGGGGTGTGA 285
DQ ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DQ 7079 GGTATCTTTGTTGACTTTCTGTCTCTGTTGATCTCTAAATGTTGACAGTGGGGTGTGA 7020
```

```

QY 286 AGTCTCCAGATATTATGTTGGAGTCTAAGTCTCTTTGTAGTCTCTAGGAGCTTCT 345
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 7019 AGTCTCCATATTATTGTTGGAGTCTAAGTCTCTTTGTAGTCTCTAGGAGCTTCT 6940
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 346 TTATGAATCTGGTCTCTCTATTGGTGCATATATATTAGGAGTCTTCTTCTT 405
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6959 TTATGAATCTGGTCTCTCTATTGGTGCATATATATTAGGAGTCTTCTTCTT 6940
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 406 GTTGAATTGATCCCTTTTAGCATATATAGTGGCCTCTTTCTCTCTTCTTCTT 465
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6899 GTTGAATTGATCCCTTTTAGCATATATAGTGGCCTCTTTCTCTCTTCTTCTT 6940
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 466 GTTAAAGTCTGTTTATACAGAGAGTTGGATTGCAA 501
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 6839 GTTAAAGTCTGTTTATACAGAGAGTTGGATTGCAA 6804
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
US-10-225-810-26/c
; Sequence 26, Application US/10225810
; Publication No. US20030157512A1
; GENERAL INFORMATION:
; APPLICANT: Birmingham, Jr., John R.
; TITLE OF INVENTION: Tramodorsins and Methods of Using Tramodorsin
; FILE REFERENCE: McLaugh-07165
; CURRENT APPLICATION NUMBER: US/10/225,810
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 26
; LENGTH: 250000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (31774)..(33774)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (42953)..(43052)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45557)..(45656)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (48203)..(48302)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (49551)..(49650)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (51561)..(51660)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (52722)..(52821)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (53864)..(53963)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (55290)..(55389)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (56674)..(56773)

```

```

; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (57879)..(57978)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (78952)..(79051)
; OTHER INFORMATION: n is a, c, g, or t
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (85316)..(85415)
; OTHER INFORMATION: n is a, c, g, or t
; US-10-225-810-26

Query Match 36.1%; Score 431.2; DB 12; Length 250000;
Best Local Similarity 91.7%; Pred. No. 1.4e-106;
Matches 473; Conservative 0; Mismatches 28; Indels 15; Gaps 1;

QY 1 GGTTTTGGTGGATTCTTAACTGAGTTCTAGTTGATTGCTGAGGAGTGCCTTAGTCCACTATGT 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 124172 GGTTTTGGTGGATTCTTAACTGAGTTCTAGTTGATTGCTGAGGAGTGCCTTAGTCCACTATGT 124113
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 61 AGTTTGTGTAAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCCTTAGTCCACTATGT 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 124112 AGTTTGTGTAAATTTCTGTTCTTTTACATTTGCTGAGGAGTGCCTTAGTCCACTATGT 124053
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 121 GGTCAAATTTGGAATAGTGTGGTGGTCTGASAGAAATGTATATCTGTTGATTTGG 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 124052 GGTCAAATTTGGAATAGTGGTGGTCTGASAGAAATGTATATCTGTTGATTTGG 123993
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 GGTTTAGAGTCTCTGTAGATGTCTATTAGGTCACATTTGCTGAGGAGTGCCTTAGTCCACTATGT 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123992 GGTTTAGAGTCTCTGTAGATGTCTATTAGGTCACATTTGCTGAGGAGTGCCTTAGTCCACTATGT 123933
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 GGA-----TCTGTCTCTGTTGATCIGTCTAATATTGACAGTGGGGTGTTGA 285
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123932 GGATATCCTTGTAACTTCTGTCTCAITGATCTGTCTAATGTTGACAGTGGGGTGTTAA 123873
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 286 AGTCTCCAGATATTATTGTTGGAGTCTAAGTCTCTTTGAGGTCCTTAGGAGTCTTCT 345
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123872 AGTCTCCAGATATTATTGTTGGAGTCTAAGTCTCTTTGAGGTCCTTAGGAGTCTTCT 123813
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 346 TTATGAATCTGGTCTCTCTATTGGTGCATATATATTAGGAGTCTTCTTCTTCTT 405
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123812 TTATGAATCTGGTCTCTCTATTGGTGCATATATATTAGGAGTCTTCTTCTTCTT 123753
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 406 GTTGAATTGATCCCTTTTAGCATATATAGTGGCCTCTTTCTCTCTTCTTCTTCTT 465
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123752 GTTGAATTGATCCCTTTTAGCATATATAGTGGCCTCTTTCTCTCTTCTTCTTCTT 123693
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 466 GTTAAAGTCTGTTTATACAGAGAGTTGGATTGCAA 501
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 123692 GTTAAAGTCTGTTTATACAGAGAGTTGGATTGCAA 123657
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
US-09-764-891-7392/c
; Sequence 7392, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7392
; LENGTH: 5446
; TYPE: DNA
; ORGANISM: Homo sapiens

```


JS-39-764-892-7392

Query Match 85.7%; Score 429.6; DB 11; Length 5446;
Best Local Similarity 91.5%; Pred. No. 7.5e-107;
Matches 472; Conservative 0; Mismatches 29; Index 15;

QY	1	GGTTTTGGCGAGATTTCTTAATCCTGAGTTCCTAGTTTGATTCGACTGTGGCCTGAGAGAC	60
DB			
DB	2091	GGCTTTGAGTGAGATTCTTAATCCTGAGTTCCTAGTTTGATTCGACTGTGGTCTGAGAGAC	2032
QY	61	AGTTTGTGTAATTTCTGTCTCTTTTACATTTTGGCTGAGGAGTGGCTTTAGTTCCAACTATGT	120
DB			
DB	2031	AGTTTGTATATATTTCTGTCTCTTTTACATTTTGGCTGAGGAGAGCTTTTACTTCCAACTATGT	1972
QY	121	GGTCAATTTTGGAAATAGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT	180
DB			
DB	1971	GGTCAATTTTGGAAATAGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGT	1912
QY	181	GGTTTAGAGTTCCTGTAGATGTCTATTAGGTCGACTTTGGTGGAGAGCTGAGTTCAATTCCT	240
DB			
DB	1911	GGTGGAGAGTTCTGTAGATGTCTATTAGGTCGCTTTGGTGGAGAGCTGAGTTCAATTCCT	1852
QY	241	GG-----AAGTGTCTTGTGATCTGTCTAATATTGACAGTGGGGTGTGTGA	285
DB			
DB	1951	GGGTATCCTTGTTGACTTTCTGTCTCGTTGATCTGTCTAATATTGACAGTGGGGTGTGTGA	1792
QY	286	AGTCTCCAGATTTATTGTGTGGAGAGTCTTAAGTCTCTTTGTAGGCTCTCTAGGAGCTTGT	345
DB			
DB	1791	AGTCTCCCATTTATTAATGTGTGGAGTCTTAAGTCTCTTTGTAGGCTCACTGAGGACTTSGT	1732
QY	346	TTATGAATCTGGGTGCTCTGTATTTGGTGCATATATATTTAGGATAGTTAGCTCTCTCT	405
DB			
DB	1731	TTATGAATCTGGGTGCTCTGTATTTGGTGCATATAATTTAGGATAGTTAGCTCTCTCT	1672
QY	406	GTTGAATTTGATCCCTTTTAGCATTTATATGATGGGCTTCTTTTGTCTCTTTTATCTTTGTTG	465
DB			
DB	1671	GTTGAATTTGATCCCTTTTACCATTTATGTAATGGGCTTCTTTTGTCTCTTTTATCTTTGTTG	1612
QY	466	GTTTAAAGTCTGTTTTTATCAGAGAGTTGGATTTGGAA	501
DB			
DB	1611	GTTTAAAGTCTGTTTTTATCAGAGAGCTTGGATTTGGAA	1576

```

RESULT 11
US-09-764-877-3755
; Sequence 3755, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PCC05
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3755
; LENGTH: 5671
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-877-3755

```

```

Query Match      85.7%; Score 429.6; DB 10; Length 567;
Best Local Similarity 91.5%; Pred. No. 7.6e-107;
Matches 472; Conservative 3; Mismatches 29; Indels 15; Gaps 1;

CY 1 GGTTTTCGGTGAGTTTCCTAATCCCTGAGTCTCTAGTTTCATTGCGTCTGCGCCGAGAGAC 60
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cb 3318 GGTTCGAGTGAGTTTCCTAATCCCTGAGTCTCTAGTTTCATTGCGTCTGCGCCGAGAGAC 3377

CY 61 AGTTTGTTGTGAATTTTCGTCTCTTTTACATTTCGTGAGGAGTGGCTTTTAGTTCCTCACTAATGT 120
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3378 AGTTTGTTTAAATTTTCGTCTCTTTTACATTTCGTGAGGAGTGGCTTTTAGTTCCTCACTAATGT 3437

```

```

CY 121 GGTC AATTTTGG AATAGG TGCGGTGGTGCAGAGAAATGTATATCTGTGATTGG 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3438 GGCC AATTTTGG AGTAAAGTSCAGTGTGGTGCAGAGAAATGTATATCTGTGATTGG 3499
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

CY 181 GGTTCAGAGTCTGTAGATGCTATATTAGGTCACACTGGGTCAGAGCTGAGTTCAGTTCCT 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3498 GGTGAGAGTCTGTAGATGCTATATTAGGTCACACTGGGTCAGAGCTGAGTTCAGTTCCT 3559
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

CY 241 SGA-----TCTGCTCTTGTCTGATCTGCTCTAATATTTGACAGTGGGGTCTTGA 285
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3558 SGATATCCTTGTCTAACCTTCCTGCTCTCATTTGATCTGCTCTAATGTTGACAGTGGGGTCTTAA 3619
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

CY 286 AGTCTCCAGCTATTATTGTGTGGGAGCTCAAGTCTCTCTTTGTAGGTCTCTAGGAGACTTGCT 345
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3618 AGTCTCCAGCTATTATTGTGTGGGAGCTCAAGTCTCTCTTTGTAGGTCTCTAGGAGACTTGCT 3679
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

CY 346 TTATGAATCTGGGTGCTCTCTGCTATTTGGGTGCATATATATTTAGATAGTATAGCTCTCTCT 405
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3678 TTATGAATCTGGGTGCTCTCTATATTTGGGTGCATATATATTTAGATAGTATAGCTCTCTCT 3739
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

CY 406 GTTGAATCTGATCTCCCTTTAGCATATATATGATGGCTCTCTTTGTCTCTCTTTTGATCTTTGTG 465
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3738 GTTGAATCTGATCTCCCTTTACCATATGTAATGGCTCTCTTTGTCTCTCTTTTGATCTTTGTG 3799
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

CY 466 GTTTAAAGCTGTCTTTATCCAGAGAGTTGGATTGCAA 501
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3798 GTTTAAAGCTGTCTTTATCCAGAGACTAGGATTGCAA 5833
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-09-764-891-10041
; Sequence 10041, Application US/09764891
; Publication No. US20030377808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P0006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10041
; LENGTH: 20746
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-891-10041

```

```

RESULT 12
US-09-764-891-10041
; Sequence 10041, Application US/09764891
; Publication No. US20030377808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P0006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1023:
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 10041
; LENGTH: 20746
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-891-10041

```

Query Match	85.7%	Score 429.6	DB 11	Length 20746	
Best Local Similarity	91.5%	Pred. No. 23e-106			
Matches 472	Conservative	0	Mismatches 29	Indels 15	Gaps 17
QY	1	GGTTTGGGTGAGTTCTCTAAATCCTGAGTTCCTAGTTTGATTGCATGTGTGGCCTGAGAGAC	60		
DB	18494	GGTTTGGGTGAGTCTCTCTAAATCCTGAGTTCCTAGTGTGATTGCATGTGTGGTCTGAGAGAC	18553		
QY	61	AGTTTGTGTTAAATTCCTGTCTCTTTACATTTGCTCAGGAGTGTCTTAGTTCCAACTATGT	120		
DB	18554	AGTTTGTGTTAAATTCCTGTCTCTTTACATTTGCTCAGGAGTGTCTTAGTTCCAACTATGT	18613		
QY	121	GGTCAATTTTGGAAATAGGTGTGGTGTGTGCTGAGAGAAGATGTATATTTCTGTTGATTTGG	180		
DB	18614	GGTCAATTTTGGAAATAGGTGTGGTGTGTGCTGAGAGAAGATGTATATTTCTGTTTATTTGG	18673		
QY	181	GGTTTAGAGTTCGTAGATGTCCTATTAGTCCACCTTGGTGCAGAGCTGAGTTCAGTTCCCT	240		
DB	18674	GGTGGAGAGTTCGTAGATGTCCTATTAGTCTCGTGGTGGTGCAGCTGCTGAGTTCAGTTCCCT	18733		
QY	241	GGG-----TCTGTCTTGTGATCTGTCTTAATATTTGACAGTGGGGTGTGTGA	285		
DB	18734	GGATATCCTTGTTAAATGTTCTGTCTCGTTGATCTGTCTTAATGTTGACAGTGGGGTGTGTGA	18793		

```
QY 286 AGTCTCCCACTATTATTGTGGGAGCTCTAAGTCTCTTTGTAGCTCTCTAGGACTTGGT 345
DB 18794 AGTCTCCCACTATTATTGTGGGAGCTCTAAGTCTCTTTGTAGCTCTCTAGGACTTGGT 18853
QY 346 TTATGAATCTGGTCTCTGTATGGGTGCATATATATTAGCATAGTTAGCTCTTCTT 405
DB 18954 TTATGAATCTGGTCTCTGTATGGGTGCATATATATTAGCATAGTTAGCTCTTCTT 18913
QY 406 GTTGAATGATCCCTTAGCATATATATGATGGCCCTCTTTGTCTCTTTTGTATCTTTG 465
DB 18914 GTTGAATGATCCCTTAGCATATATATGATGGCTCTTTTGTCTCTTTTGTATCTTTG 18973
QY 466 GTTTAAAGTCTGTTTATCATCAGAGACTAGGATTGCAA 501
DB 18974 GCTTAAAGTCTGTTTATCATCAGAGACTAGGATTGCAA 19039

RESULT 13
US-09-818-512-3
; Sequence 3, Application US/39819512
; Patent No. US20020142416A1
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001192
; CURRENT APPLICATION NUMBER: US/09/818,512
; CURRENT FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 116592
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(116592)
; OTHER INFORMATION: n = A,T,C or G
US-09-818-512-3

Query Match 85.4%; Score 428; DB 10; Length 116592;
Best Local Similarity 91.3%; Pred. No. 7.6e-106;
Matches 471; Conservative 0; Mismatches 30; Indels 15; Gaps 1;

QY 1 GGTCTTGGGTGAGTTTCTTAATCTGAGTTCTAGTTTGAATTCGACTGTGGCCCTGAGAGAC 60
DB 51828 GGTCTTGGGTGAGATTCTTAATCTGAGTTCTAGTTTGAATTCGACTGTGGCCCTGAGAGAC 51887
QY 61 AGTTTGTGTAAATTTCTCTTTTACATTTCTGATGAGAGAGTGGTTAGTTCCAACTATGT 120
DB 51888 AGTTTGTGTAAATTTCTCTTTTACATTTCTGATGAGAGAGTGGTTAGTTCCAACTATGT 51947
QY 121 GGTCAATTTTGGAACTAGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
DB 51948 GGTCAATTTTGGAACTAGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 52007
QY 181 GGTCTAGAGTTCTGTAGATGTCTATTAGTTCGACTTGGTGCAGAGCTGAGTTCACTTGGT 240
DB 51948 GGTCAATTTTGGAACTAGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 52007
QY 241 GG-----ATCTGCTCTTGTGATCTGCTAAATATTGACAGTGGGGTGTGGA 285
DB 52068 GGTATCTCTTGTGACTTTCTGCTCTGTTGATCTGCTAAATATTGACAGTGGGGTGTGGA 52127
QY 286 AGTCTCCCACTATTATTGTGGGAGTCTAAGTCTCTTTGTAGGCTCTCTAGGACTTGGT 345
DB 52068 GGTATCTCTTGTGACTTTCTGCTCTGTTGATCTGCTAAATATTGACAGTGGGGTGTGGA 52127
QY 346 TTATGAATCTGGTCTCTGTATGGGTGCATATATATTAGCATAGTTAGCTCTTCTT 405
DB 52068 GGTATCTCTTGTGACTTTCTGCTCTGTTGATCTGCTAAATATTGACAGTGGGGTGTGGA 52127
QY 406 GTTGAATGATCCCTTAGCATATATATGATGGCCCTCTTTGTCTCTTTTGTATCTTTG 465
DB 52128 AGTCTCCCACTATTATTGTGGGAGTCTAAGTCTCTTTGTAGGCTCTCTAGGACTTGGT 52187
QY 466 GTTTAAAGTCTGTTTATCATCAGAGACTAGGATTGCAA 501
DB 52188 GTTTAAAGTCTGTTTATCATCAGAGACTAGGATTGCAA 52347
```

```
QY 466 GTTGAATGATCCCTTAGCATATATATGATGGCCCTCTTTGTCTCTTTTGTATCTTTG 465
DB 52248 GTTGAATGATCCCTTAGCATATATATGATGGCCCTCTTTGTCTCTTTTGTATCTTTG 52307
QY 466 GTTTAAAGTCTGTTTATCATCAGAGACTAGGATTGCAA 501
DB 52308 GTTTAAAGTCTGTTTATCATCAGAGACTAGGATTGCAA 52343

RESULT 14
US-10-354-065-3
; Sequence 3, Application US/10354065
; Publication No. US20030138837A1
; GENERAL INFORMATION:
; APPLICANT: BEASLEY, Ellen et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001192DIV
; CURRENT APPLICATION NUMBER: US/10/354,065
; CURRENT FILING DATE: 2003-01-30
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 116592
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(116592)
; OTHER INFORMATION: n = A,T,C or G
US-10-354-065-3

Query Match 85.4%; Score 428; DB 12; Length 116592;
Best Local Similarity 91.3%; Pred. No. 7.6e-106;
Matches 471; Conservative 0; Mismatches 30; Indels 15; Gaps 1;

QY 1 GGTCTTGGGTGAGTTTCTTAATCTGAGTTCTAGTTTGAATTCGACTGTGGCCCTGAGAGAC 60
DB 51828 GGTCTTGGGTGAGATTCTTAATCTGAGTTCTAGTTTGAATTCGACTGTGGCCCTGAGAGAC 51887
QY 61 AGTTTGTGTAAATTTCTCTTTTACATTTCTGATGAGAGAGTGGTTAGTTCCAACTATGT 120
DB 51888 AGTTTGTGTAAATTTCTCTTTTACATTTCTGATGAGAGAGTGGTTAGTTCCAACTATGT 51947
QY 121 GGTCAATTTTGGAACTAGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
DB 51948 GGTCAATTTTGGAACTAGGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 52007
QY 181 GGTCTAGAGTTCTGTAGATGTCTATTAGTTCGACTTGGTGCAGAGCTGAGTTCACTTGGT 240
DB 52008 GGTGAGAGTTCTGTAGATGTCTATTAGTTCGACTTGGTGGTGGTGGTGGTGGTGGTGG 52067
QY 241 GG-----ATCTGCTCTTGTGATCTGCTAAATATTGACAGTGGGGTGTGGA 285
DB 52068 GGTATCTCTTGTGACTTTCTGCTCTGTTGATCTGCTAAATATTGACAGTGGGGTGTGGA 52127
QY 286 AGTCTCCCACTATTATTGTGGGAGTCTAAGTCTCTTTGTAGGCTCTCTAGGACTTGGT 345
DB 52128 AGTCTCCCACTATTATTGTGGGAGTCTAAGTCTCTTTGTAGGCTCTCTAGGACTTGGT 52187
QY 346 TTATGAATCTGGTCTCTGTATGGGTGCATATATATTAGCATAGTTAGCTCTTCTT 405
DB 52188 TTATGAATCTGGTCTCTGTATGGGTGCATATATATTAGCATAGTTAGCTCTTCTT 52247
QY 406 GTTGAATGATCCCTTAGCATATATATGATGGCCCTCTTTGTCTCTTTTGTATCTTTG 465
DB 52248 GTTGAATGATCCCTTAGCATATATATGATGGCCCTCTTTGTCTCTTTTGTATCTTTG 52307
QY 466 GTTTAAAGTCTGTTTATCATCAGAGACTAGGATTGCAA 501
DB 52308 GTTTAAAGTCTGTTTATCATCAGAGACTAGGATTGCAA 52343
```

```
RESULT 15
US-10-027-983-11/c
; Sequence 11, Application US/10027983
; Publication No. US2003013936CA1
; GENERAL INFORMATION:
; APPLICANT: Kenneth W. Doobie
; APPLICANT: Mark P. Roach
; TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION
; FILE REFERENCE: RTS-0340
; CURRENT APPLICATION NUMBER: US/10/027,983
; CURRENT FILING DATE: 2001-12-18
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 11
; LENGTH: 392000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 137743
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 137742
; OTHER INFORMATION: unknown
; NAME/KEY: misc feature
; LOCATION: 138122...138221
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: unsure
; LOCATION: 145507
; OTHER INFORMATION: unknown
; NAME/KEY: unsure
; LOCATION: 151967
; OTHER INFORMATION: unknown
; NAME/KEY: misc feature
; LOCATION: 151967...1542066
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: unsure
; LOCATION: 154217
; OTHER INFORMATION: unknown
; NAME/KEY: misc feature
; LOCATION: 164037...164136
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: 174657...174756
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: 186224...186323
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: 195242...195341
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: unsure
; LOCATION: 202703
; OTHER INFORMATION: unknown
; NAME/KEY: misc feature
; LOCATION: 202771...202870
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: 206246...215602
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: 218126...218225
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: 220360...220459
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: 222717...222816
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: 223381...224080
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
```

```
; LOCATION: (227487)...(227586)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (230157)...(230256)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (232299)...(232398)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (236552)...(236665)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (238789)...(248788)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: exon
; LOCATION: (118288)...(119101)
; OTHER INFORMATION: exon 10
; NAME/KEY: exon:intron junction
; LOCATION: (151129)...(151130)
; OTHER INFORMATION: exon 5:intron 5
; NAME/KEY: exon:intron junction
; LOCATION: (299248)...(299249)
; OTHER INFORMATION: exon 9:intron 9
; NAME/KEY: exon:intron junction
; LOCATION: (348578)...(348579)
; OTHER INFORMATION: exon 10:intron 10
; NAME/KEY: intron
; LOCATION: (348579)...(381838)
; OTHER INFORMATION: intron 10
; NAME/KEY: intron:exon junction
; LOCATION: (386185)...(386186)
; OTHER INFORMATION: intron 10:exon 12
US-10-027-983-11

Query Match 95.4% Score 428; DB 12; Length 392000;
Best Local Similarity 91.3%; Pred. No. 1.3e-105;
Matches 473; Conservative 0; Mismatches 30; Indels 15; Gaps 1;

QY 1 GGTTCGCGTGAGTTTCCTTAATCCTGAGTCTAGTTGATGCGACTGTCGCTGAGAGAC 60
Db 318449 GGTTCGAGTGAGTTTCCTTAATCCTGAGTCTAGTTGATGCGACTGTCGCTGAGAGAT 318390
QY 61 AGTTGTTGTAANTTCCTGTTCTTTTACATTTGCTGAGGAGTGCCTTTAGTCCAACTATGT 120
Db 318389 AGTTGTTAATAATTCCTGTTCTTTTACATTTGCTGAGGAGTGTTCCTTACCTCCAACTATGT 318330
QY 121 GGTCAATTTGGGAATAGGTGCTGCTGAGAGAGTGTGCTGAGAGAGTGTATATTCCTGATTGG 180
Db 318329 GGTCAATTTGGGATTAAGTCGGCGTGGTCTGAGAGAGTGTATATTCCTGATTGG 318270
QY 181 GGTTCAGAGTCTGTAGATGCTCTATTAGTCCACTTGTGTCAGAGCTGAGTTCCATTCCT 240
Db 318269 GGTGGAGAGTCTGTAGATGCTCTATTAGTCTGCTTGTGTCAGAGCTGAGTTCCATTCCT 318210
QY 241 GGA-----TCTGCTTTGTTGATCTGTCTAATAATTGACAGTGGGTGTTGA 285
Db 318209 GGATATCCTTGTTAATTTCTGCTCTGTTGATCTGTCTAATGTTGACAGTGGGATGTTAA 318150
QY 286 AGTCTCCAGTATTATTGTTGGGAGTCTAAGTCTCTTTGTAGGCTCTAGGACTTGGT 345
Db 318149 AGTCTCCGATTATTATTGTTGGGAGTCTAAGTCTCTTTGTAGGTCATTAAGGACTTGGT 318090
QY 346 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTACTCTTTT 405
Db 318089 TTATGAATCTGGGTGCTCCTGTATTGGGTGCATATATATTTAGGATAGTACTCTTTT 318030
QY 406 GTTGAATTCATCCCTTTAGCATTATATGATGGCCTCTTTTGTCTCTTTTGTCTTTTGTG 465
Db 318029 GTTGAATTCATCCCTTTACTATTATGTAAATGGCCTCTTTTGTCTCTTTTGTCTTTG 317970
QY 466 GTTTAAAGTCTGTTTTTATCAGAGAGTTGGATTGCAA 501
Db 317969 GTTTAAAGTCTGTTTTTAAACAGAGACTAGGATTGCAA 317934
```

Search completed: October 24, 2003, 18:34:50
CCP time : 130.415 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 24, 2003, 18:12:56 ; Search time 1046.77 Seconds
(without alignments)
11633.309 Million cell updates/sec

Title: US-09-830-902-1_COPY_50000_50500

Perfect score: 501
Sequence: 1 ggttttgcgaggtttctctta.....atcagagagattgagatcgaa 501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estha:
2: em_esthum:
3: em_estin:
4: em_estmu:
5: em_estov:
6: em_estpi:
7: em_estro:
8: em_hic:
9: gb_est1:
10: gb_est2:
11: gb_hic:
12: gb_est3:
13: gb_est4:
14: gb_est5:
15: em_estfun:
16: em_estom:
17: em_gss_hum:
18: em_gss_inv:
19: em_gss_pin:
20: em_gss_vrt:
21: em_gss_fur:
22: em_gss_mam:
23: em_gss_mus:
24: em_gss_pro:
25: em_gss_rtd:
26: em_gss_phg:
27: em_gss_vrl:
28: gb_gss1:
29: gb_gss2:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	426.4	85.1	753	13	BU959549
C 2	426.4	85.1	1075	21	BC012129
C 3	424.8	84.8	880	10	BG619398
C 4	423.8	84.6	2495	11	BC037952

C 5	423.2	84.5	609	2	HSX074999
C 6	422.4	84.3	579	3	AL119263
C 7	422.4	84.3	579	3	AL133777
C 8	421.2	84.1	651	28	AC385776
C 9	420.6	84.0	778	13	BU567312
C 10	420	83.8	1100	12	BM473445
C 11	420	83.8	3314	28	AF101615
C 12	418.4	83.5	729	29	BZ772880
C 13	415.4	82.9	604	12	B1018878
C 14	415.4	82.9	609	12	B1021312
C 15	415.4	82.9	626	12	B1019813
C 16	415.2	82.9	750	3	AU122802
C 17	414.2	82.7	701	2	HSM087935
C 18	413.6	82.6	777	14	CB312020
C 19	412.8	82.4	688	10	BG572005
C 20	412	82.2	1037	29	BZ601843
C 21	410.4	81.9	609	10	AW860981
C 22	410.4	81.9	829	29	BZ774229
C 23	410	81.8	712	28	AQ055327
C 24	406.4	81.1	799	14	CD511598
C 25	404.8	80.8	651	29	AG019019
C 26	404.8	80.8	804	9	AV711080
C 27	402.2	80.3	698	10	BE173627
C 28	400	79.8	574	2	HSM079550
C 29	399.8	79.8	707	23	AQ470254
C 30	398.8	79.6	497	28	AQ428674
C 31	395.6	79.0	687	10	B3173465
C 32	395	78.8	554	10	BZ676389
C 33	393.4	78.5	829	10	BG537539
C 34	392.6	78.4	615	10	AW856417
C 35	392.6	78.4	654	9	AV730783
C 36	392.2	78.3	658	29	AG034925
C 37	392.2	78.3	732	29	BZ612233
C 38	390	77.8	709	13	BQ694643
C 39	389.6	77.8	663	29	AG119932
C 40	388.2	77.5	668	23	AQ422387
C 41	388.2	77.5	701	29	AG116167
C 42	385	76.8	557	29	AG019827
C 43	384.6	76.8	703	28	AZ917428
C 44	382	76.2	623	28	AQ555093
C 45	381.8	76.2	909	29	BZ598445

ALIGNMENTS

RESULT 1
BU959549/C
LOCUS
DEFINITION AGENCOURT_10623586 NIH_MGC_127 Homo sapiens cDNA clone
IMAGE:6737882 5', mRNA sequence.
ACCESSION BU959549
VERSION BU959549.1 GI:24189121
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
BU959549 753 bp linear EST 21-OCT-2002
AGENCOURT_10623586 NIH_MGC_127 Homo sapiens cDNA clone
IMAGE:6737882 5', mRNA sequence.
BU959549.1 GI:24189121
EST.
Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Accession to 753
NIH-MGC <http://mgi.nlm.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: NCI
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LJNL)
ENA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LJNL at:
<http://image.llnl.gov>
Plate: LCM368 row: k column: 01
High quality sequence stop: 526.


```
RESULT 11
AF101615/c
LOCUS AF101615 Human Homo sapiens genomic clone pTWS233.20, genomic
DEFINITION AF101615 3314 bp DNA linear GSS 06-NOV-2000
survey sequence.
ACCESSION AF101615
VERSION AF101615.1 GI:4249718
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3314)
AUTHORS Bepier, G., O'Brian, K.C., Kim, Y.C., Schreiber, G. and Pitterle, D.M.
TITLE A 1.4-Mb high-resolution physical map and contig of chromosome
segment 1p15.5 and genes in the LCH1A metastasis suppressor
region
JOURNAL Genomics 55 (2), 164-175 (1999)
MEDLINE 99334294
PUBMED 9933563
COMMENT Contact: Bepier G
Medicine and Radiology
Duke University Medical Center
Box 2610, MSRB, Room 117, Durham, NC 27710, USA
part of a 1.4 megabase contig including the LCH1A metastasis
suppressor region B1: B
Class: unknown.
FEATURES
source
location/Qualifiers
1..3314
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/map="11p15.5"
/clone="pTWS233.20"
/clone_lib="Human"
BASE COUNT 1300 a 743 c 670 g 601 t
Query Match 83.8%; Score 420; DB 28; Length 3314;
Best Local Similarity 90.3%; Pred. No. 4.9e-63;
Matches 466; Conservative 0; Mismatches 35; Indels 15; Gaps 1;
QY 1 GGTTCGGTGAGTTCCTTAATCCTGAGTCTAGTTTGATTCGACTGGCCGAGAGAC 60
DB 2618 GGTTCGAGTTCCTTAATCCTGAGTCTAGTTTGATTCGACTGGCTGAGAGAC 2559
QY 61 AGTTGTTGTAATTTCTGTTCTTTTACATTCCTGAGGAGTCTCTTTAGTCCCACTATGT 120
DB 2558 AGTTGTTGTAATTTCTGTTCTTTTACATTCCTGAGGAGTCTCTTTAGTCCCACTATGT 2499
QY 121 GGTCAATTTTGGAAATAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 180
DB 2498 GGTCAATTTTGGAAATAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2439
QY 181 GGTTCAGATTCCTGTAGATGCTATATTAGTCCACATTCGTCGAGAGCTGATTCCT 240
DB 2438 GGTTCAGATTCCTGTAGATGCTATATTAGTCCACATTCGTCGAGAGCTGATTCCT 2379
QY 241 GGA-----TCTGTCTTTGTTGATCTGTCTAATATTGACAGTGGGGGTGTTGA 285
DB 2378 GGATATCCCTTGTAACTTTCTGTCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 2319
QY 286 AGTCTCCCACTATTATTGTGGGAGTCTAAGTCTCTTTGTTAGGTCTCTAGGGACTTCT 345
DB 2318 AGTCTCCCACTATTATTGTGGGAGTCTAAGTCTCTTTGTTAGGTCTCTAGGGACTTCT 2259
QY 346 TTATGAATCTGGGTGCTCCGTGATTTGGGTGCATATATATTAGGATAGTCTCTCTCT 405
DB 2258 TTATGAATCTGGGTGCTCCGTGATTTGGGTGCATATATATTAGGATAGTCTCTCTCT 2199
QY 406 GTTGAATTGATCCCTTTAGGATTTATATGATGGCCCTCTTTGCTCTCTCTCTCTCTCT 465
```

```
DB 2198 CTTGAATAGATCCCTTTGCCATTATGTAATGGCCTTCTTTGTCTCTTTTGATCTTTGTTG 2139
QY 466 GTTTAAAGTCTGTTTATCAGAGAGATTGGATTGCAA 501
DB 2138 GTTTAAAGTCTGTTTATCAGAGACTAGGATTGCAA 2103
RESULT 12
BZ772880
LOCUS BZ772880 729 bp DNA linear GSS 13-MAR-2003
DEFINITION BZ772880 Homo sapiens genomic, genomic survey
sequence.
ACCESSION BZ772880
VERSION BZ772880.1 GI:28946564
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 729)
AUTHORS Cook, L., Delehaunty, K., Fewell, G., Fulton, L., Magrini, V., Vardis, E.,
Miner, T., Nash, W., Williams, D. and Wilson, R.K.
TITLE Homo sapiens Fosmid End Reads
JOURNAL Unpublished
COMMENT Contact: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@watson.wustl.edu
Plate: mv56 row: c column: 10
Class: fosmid ends
High quality sequence start: 38
High quality sequence stop: 550.
FEATURES
location/Qualifiers
1..729
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone_lib="HPOSMIDC07"
/note="Vector: pCC1fos; Site 1: Eco721; Human whole
genome fosmid library was prepared at Washington
University Genome Sequencing Center. DNA was sheared for
blunt-ended ligation into pCC1fos inducible vector. DNA
was ordered from Coriell Cell Repository's DNA
polymorphism discovery resource."
BASE COUNT 145 a 120 c 172 g 289 t
ORIGIN
Query Match 83.5%; Score 418.4; DB 29; Length 729;
Best Local Similarity 90.1%; Pred. No. 1.1e-62;
Matches 465; Conservative 0; Mismatches 36; Indels 15; Gaps 1;
```

DE 504 AGTCTCCCATTAATAGTGTGGAGTCTAAGTCTCTCTTTGTAGTCACTAGGACTCGCT 563
CY 346 TTATGAATCTGGTCTCTCTGATTTGGGTGCAATATATATTTAGCAATAGTACTCTCTTT 405
DB 564 TTATGAATCTGGTCTCTCTGATTTGGGTGCAATATATATTTAGCAATAGTACTCTCTTT 623
CY 406 GTTGAATTGATCCCTTTAGCAATATATGATGGCCCTCTCTTTGATCTCTCTTTGATCTTTG 465
DB 624 GTTGAATTGATCCCTTTAGCAATATATGATGGCCCTCTCTTTGATCTCTCTTTGATCTTTG 683
CY 466 GTTTAAAGTCTGTTTATACAGAGAGTTGGATTGCAA 501
DB 684 GTTTAAAGTCTGTTTATACAGAGAGTTGGATTGCAA 719

RESULT 13
BI018878/c
LOCUS BI018878 604 bp mRNA linear EST 14-JUN-2001
DEFINITION IL3-MT0267-050101-435-B02 MT0267 Homo sapiens cDNA, mRNA sequence.
ACCESSION BI018878
VERSION BI018878.1 GI:14425508
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 604)
AUTHORS Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.P., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,J.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL3&at2=IL3-MT0267-
050101-435-B02&t3=2001-01-05&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 603.
Location/Qualifiers
1..604
/organism="Homo sapiens"
/roi_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="MT0267"
/note="Organ: marrow; Vector: puc18; Site: 11; Smal: Site 2;
Smal: A mini-library was made by cloning products derived
from ORFESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 253 a 142 c 92 g 117 t
ORIGIN
Query Match 82.9%; Score 415.4; DB 12; Length 604;
Best Local Similarity 90.9%; Pred. No. 3.8e-62;
Matches 470; Conservative 0; Mismatches 31; Indels 16; Gaps 2;

CY 1 GGTTTTGGCTGAGTTCTTAATCTGAGTTCTAGTTTGTAGTCACTGTGGCCTTGAGAGAC 60
DB 596 GGTTTTGGCTGAGTTCTTAATCTGAGTTCTAGTTTGTAGTCACTGTGGTCTGAGAGAT 537
CY 61 AGTTTGTGTAATTTCTGTTTCTTTTACATTTGGTGAGGAGTGGTTTAGTTCACAACTATGT 120
DB 536 AGTTTGTGTAATTTCTGTTTCTTTTACATTTGGTGAGGAGAACTTTACTTCCAACTATGT 477
CY 121 GGTCAATTTTGGAAATAGGTGTGGTGTGGTGTG-AGAAGAAATGTATATCTGTGATTTG 179
DB 476 GGTCAATTTTGGAAATAGGTGTGGTGTGGTGTG-AAAAAATGTATATCTGTGATTTG 417
CY 180 GGGTTTAGAGTCTCTGTAGATGTCTATTAGGTCCACTTTGGTGAGAGCTGAGTTCAGTTCC 239
DB 416 GGTGGAGAGTCTCTGTAGATGTCTATTAGGTCCGCTTGGTGAGAGCTGAGTTCAAATTC 357
CY 240 TGG-----ATCTGCTCTTGTGTGATCTGCTTAATTTGACAGTGGGGTGTG 284
DB 256 TGGTATCTCTTGTAACTTTCTGCTCTGTTGATCTGTCTAATGTGACATTTGGGGTGTGTA 297
CY 285 AAGTCTCCAGTATATTGTGTGGGAGTCTAAGTCTCTTTGTAGGTTCTCTAGGACTTGC 344
DB 296 AAGTCTCCCATTTAATGTGTGGGAGTCTAAGTCTCTTTGTAGGTTCACTCAGGACTTGC 237
CY 345 TTATGAATCTGGTGTCTCTGTATTGGGTGCATATATATTAGGATAGTTAGTCTCTCT 404
DB 236 TTATGAATCTGGTGTCTCTGTATTGGGTGCATATATATTAGGATAGTTAGTCTCTCT 177
CY 405 TGTTGAATTGATCCCTTTAGCAATATATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTT 464
DB 176 TGTTGAATTGATCCCTTTAGCAATATATGATGGCCTTCTTTGTCTCTTTTGATCTTTGTT 117
CY 465 GGTAAAGTCTGTTTATACAGAGAGTTGGATTGCAA 501
DB 116 GGTAAAGTCTGTTTATACAGAGAGTGGATTGCAA 80

RESULT 14
BI021312/c
LOCUS BI021312 609 bp mRNA linear EST 14-JUN-2001
DEFINITION IL3-MT0267-150101-448-C10 MT0267 Homo sapiens cDNA, mRNA sequence.
ACCESSION BI021312
VERSION BI021312.1 GI:14427942
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 609)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.P., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,J.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL3&at2=IL3-MT0267-
150101-448-C10&t3=2001-01-15&t4=1)
Seq primer: puc 18 forward


```
High quality sequence stop: 609.
Location/Qualifiers
Source 1..609
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="MT0267"
Note="Organ: marrow; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 255 a 143 c 93 g 118 t
ORIGIN
Query Match 82.9%; Score 415.4; DB 12; Length 626;
Best Local Similarity 90.9%; Pred. No. 3,7e 62;
Matches 470; Conservative 0; Mismatches 31; Indels 16; Gaps 2;
CY 1 GGTCTTGGCTGAGTTCTTAACTCCTGAGTCTCTAGTTTCATTCGACCTGGGCTGAGAGAC 60
|||
DB 596 GGTCTTGGCTGAGATTCCTAACTCCTGAGTCTCTAGTTTCATTCGACCTGGGCTGAGAGAT 530
|||
CY 61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGAGAGTCTCTTACCTCCAACTATGT 120
|||
DB 536 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGAGAGTCTCTTACCTCCAACTATGT 420
|||
CY 121 GGTCAATTTTGGATAGTGTGTGGTGTGGTCTG-AGAGAAATGTATATTCTGTTGATTG 130
|||
DB 476 GGTCAATTTTGGATAGTGTGTGGTGTGGTCTG-AGAGAAATGTATATTCTGTTGATTG 410
|||
CY 180 GGGTTTAGAGTCTCTGAGATGCTATTAGTCCACTTGGTGCAGAGCTGAGTTCAGTTC 230
|||
DB 416 GGGTGGAGAGTCTCTGAGATGCTATTAGTCCACTTGGTGCAGAGCTGAGTTCAGTTC 350
|||
CY 240 TGG-----ATCTGCTCTTGTGATCTGCTTAATATTGACAGTGGGCTGTG 284
|||
DB 356 TGGGTATCCCTGTTAACTTTCTCTGCTGATCTGCTTAATGTTGACATGGGGTGTA 290
|||
CY 285 AAGTCTCCCACTATTATTGTTGGAGTCTTAAGTCTCTTGTAGTCTCTAGGAGCTTC 344
|||
DB 296 AAGTCTCCCACTATTATTGTTGGAGTCTTAAGTCTCTTGTAGTCTCTAGGAGCTTC 230
|||
CY 345 TTTATGAATCTGGTCTCTGATATTGGTGCATATATATTAGGATAGTTCCTCT 404
|||
DB 236 TTTATGAATCTGGTCTCTGATATTGGTGCATATATATTAGGATAGTTCCTCT 170
|||
CY 405 TGTGAAATGATCCCTTTAGCATTATATGATGGCCCTTCTTGTCTCTTTGATCTT 464
|||
DB 176 TGTGAAATGATCCCTTTAGCATTATATGATGGCCCTTCTTGTCTCTTTGATCTT 110
|||
CY 465 GGTTTAAAGTCTGTTTATCAGAGAGTTCGATTCGAA 501
|||
DB 116 GGTTTAAAGTCTGTTTATCAGAGAGTTCGATTCGAA 50
|||
RESULT 15
BIO19813
LOCUS 626 bp mRNA linear EST 14-JUN-2001
DEFINITION IL3-MT0267-110101-442-F12 MT0267 Homo sapiens cDNA, mRNA sequence.
ACCESSION BIO19813
VERSION BIO19813.1 GI:14426443
KEYWORDS EST.
SOURCE Homo sapiens (human);
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo;
Homo sapiens (human);
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.C., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.C. and
Simpson,A.C.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20020663
10737800
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-Sp,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
http://www.ludwig.org.br/scripts/gethtml2.pl?fl=ID3&t2=IL3-MT0267-
110101-442-F12&t3=2001-01-11&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 609.
Location/Qualifiers
Source 1..626
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="MT0267"
Note="Organ: marrow; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT 125 a 94 c 150 g 257 t
ORIGIN
Query Match 82.9%; Score 415.4; DB 12; Length 626;
Best Local Similarity 90.9%; Pred. No. 3,7e 62;
Matches 470; Conservative 0; Mismatches 31; Indels 16; Gaps 2;
CY 1 GGTCTTGGCTGAGTTCTTAACTCCTGAGTCTCTAGTTTCATTCGACCTGGGCTGAGAGAC 60
|||
DB 14 GGTCTTGGCTGAGATTCCTAACTCCTGAGTCTCTAGTTTCATTCGACCTGGGCTGAGAGAT 73
|||
CY 61 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGAGAGTCTCTTACCTCCAACTATGT 120
|||
DB 74 AGTTTGTTGTAATTTCTGTTCTTTTACATTTGCTGAGAGAGTCTCTTACCTCCAACTATGT 133
|||
CY 121 GGTCAATTTTGGATAGTGTGTGGTGTGGTCTG-AGAGAAATGTATATTCTGTTGATTG 179
|||
DB 134 GGTCAATTTTGGATAGTGTGTGGTGTGGTCTG-AGAGAAATGTATATTCTGTTGATTG 193
|||
CY 180 GGGTTTAGAGTCTCTGAGATGCTATTAGTCCACTTGGTGCAGAGCTGAGTTCAGTTC 239
|||
DB 194 GGGTGGAGAGTCTCTGAGATGCTATTAGTCCACTTGGTGCAGAGCTGAGTTCAGTTC 253
|||
CY 240 TGG-----ATCTGCTCTTGTGATCTGCTTAATATTGACAGTGGGCTGTG 284
|||
DB 254 TGGGTATCCCTGTTAACTTTCTCTGCTGATCTGCTTAATGTTGACATGGGGTGTA 313
|||
CY 285 AAGTCTCCCACTATTATTGTTGGAGTCTTAAGTCTCTTGTAGTCTCTAGGAGCTTC 344
|||
DB 314 AAGTCTCCCACTATTATTGTTGGAGTCTTAAGTCTCTTGTAGTCTCTAGGAGCTTC 373
|||
CY 345 TTTATGAATCTGGTCTCTGATATTGGTGCATATATATTAGGATAGTTCCTCTTCT 404
|||
DB 374 TTTATGAATCTGGTCTCTGATATTGGTGCATATATATTAGGATAGTTCCTCTTCT 433
|||
CY 405 TGTGAAATGATCCCTTTAGCATTATATGATGGCCCTTCTTGTCTCTTTGATCTT 464
|||
```